Identification of essential attributes in conclusions of student reports

Aurelio López-López[1*] , Samuel González-López[2*] and Jesús Miguel García-Gorrostieta[1]

* Correspondence: allopez@inaoep. mx; samuelgonzalezlopez@gmail. com

[1]Computational Sciences Department, Instituto Nacional de Astrofísica, Óptica y Electrónica, Santa María Tonantzintla, Puebla, México
[2]Department of Information Technologies, Technological University of Nogales, Nogales, Sonora, Mexico

Abstract

This work seeks to help students in improving their first research reports, based on natural language processing techniques. We present a Conclusion model that includes three schemes: Goal Connectedness, Judgment and Speculation. These subsystems try to account for the main expected attributes in conclusions, specifically the Connectedness with the general objective of the research, the evidence of value Judgments, and the presence of Future work as a result of the student reflection after the inquiry. The article details the schemes, a validation of the approach in an annotated corpus, and a pilot test with undergraduate students. Results of a prior validation indicate that student writings indeed adhere to such attributes, especially at graduate level. Statistical results of the pilot test showed that undergraduate students in an experimental group achieved improved conclusion content when compared with the control group.

Keywords: Natural language processing, Automated text evaluation, Conclusion formulation, Goal connectedness, Reports assessment

Introduction

A student report is a document describing the student's research and main findings on a topic. Often such report is further developed into a larger student thesis. Such document requires usually the guidance of an advisor. One study focused on the perceptions of students concerning difficulties when writing the discussion section of reports (Bitchener and Basturkmen 2006). The study used in-depth interviews with supervisors and students (including L2) and found that pupils mentioned the uncertainty about what content to include and how discussion sections should be organized. This was surprising, considering the time and feedback that students received from supervisors.

This paper focuses on evaluating the conclusion section of student reports and perform a pilot test with undergraduate students. These are parts of a larger project that aims to help students to evaluate their early drafts and facilitate the review process for the academic advisor. Besides, the review time can be reduced improving the quality of feedback provided by the instructor, through allowing the reviewer focusing on the conclusions content (Debuse et al. 2008).

In a conclusion section, a discussion of the results is expected, and students are required to reflect on the whole research work. A good conclusion section should include: an analysis of compliance with the research objectives, a global response to the problem statement, a contrast between the results and the theoretical framework, areas

for further research, and an acceptance or rejection of the established hypothesis (Allen 1973). A pattern that summarizes what is expected in a conclusion section is provided by the Teaching and Learning Centre at University of New England, Australia (UNE). The pattern goes from the specific to the general, and begins with a reformulation of the problem, followed by key findings, and ending with recommendations and future work. The guide pattern is similar to the conclusion of a scientific article, but more extensive.

In the conclusion pattern, the conclusion starts by pointing to the problem solved. In the five-paragraph essay paradigm (Davis and Liss 2006), the introduction and conclusion share the main topic, namely, the subject matter of the essay. The approach is like the conclusions section, as the conclusion should be related to the general objective (considering methodological guides), in its first paragraph. In the intermediate paragraphs, the student must express his thoughts and opinions, avoiding a list of results. The Online Writing Lab at Purdue University provides an outline for writing conclusion sections, emphasizing that the conclusion must contain well-argued viewpoints and avoid inclusion of additional items that are not contained within the thesis (P.O.W Lab: Purdue Online Writing Lab 2018). Future work and recommendations included in the conclusion evidence that the student has gone beyond solving the immediate problem and can identify possible expansion and implications of the work. Currently, our work focuses on quantitative thesis in the area of computer science and nearby disciplines.

Based on the previous pattern and mentioned desirable attributes, we proposed the use of an automatic analysis of conclusions intended to obtain a first diagnostic of frequent problems in student's conclusion writings. Our first goal is the design of a model, including a methodology to evaluate the conclusion, for this purpose, we formulate this analysis in terms of three main subcomponents (schemes) that identify the following attributes of conclusions: *Goal Connectedness, Judgment,* and *Speculation.* Due to the complexity of the task, this work only focuses on the conclusions section, besides of being a key section in a thesis or project report.

A system is proposed with a central Conclusion Model, integrating the three schemes, and taking advantage of a corpus to acquire the reference knowledge, to obtain the best features and set score thresholds. After evaluation of a conclusion supplied for analysis, our system will send the result to the student, with the goal of showing him the diagnosed level reached by the conclusion. The student will be able then to improve his conclusion based on the diagnosis, before submission to the advisor.

We report the use of the three attributes to assess a corpus tagged by annotators, to validate them, once they have been implemented in a computational tool. The implementation of the model in an online application to validate the model in a real environment is the second objective of this work. The third goal is providing statistical information of correlation between the three features considered in this research. The results of a pilot test with undergraduate students of engineering are included, revealing a correlation between Goal Connectedness and Judgment characteristics. Such outcome provides evidence that students are indeed connecting their value judgments with the general objective.

Related work

Automated Writing Evaluation (AWE) of student texts, also called Automated Essay Scoring (AES), refers to the process of evaluating and scoring written text using a computer

system. Such a system builds a scoring model by extracting linguistic features (lexical, syntactic or semantic) on a specific corpus that has been annotated by humans. For this task, the researchers have been using artificial intelligence techniques such as natural language processing (NLP) and machine learning. The system can be used to directly assign a score or a quality level to a student text (Gierl et al. 2014). The use of AWE systems offers students ways to improve their writing in an automated manner, and helps to reduce review time required by academic advisors and is a complementary tool to their work.

Currently, the advances in AWE systems include the use of natural language processing technologies to perform the evaluation of texts and provide feedback to students. In this context, the system Writing Pal (WPal) offers strategy instruction and game-based practice in the writing process for developing writers. WPal assesses essay quality using a combination of computational linguistics and statistical modelling. Different linguistic properties were selected as predictors (Crossley et al. 2013). Similarly, our work seeks to assess the text attributes focusing on the conclusion section of a research report, considering three schemes to evaluate it.

In (McNamara et al. 2010), the aim was to distinguish differences between low and high scoring essays of undergraduate students. They used the Coh-Metrix tool and found that essays with a higher score reflected more sophisticated language and text complexity. In addition, using a holistic approach of quality text in (Crossley et al. 2016), the authors conducted an analysis of four features that together evidence the presence of the construct "idea generation" in student essays. Fluency, flexibility, originality, and elaboration were the elements analyzed. The corpus consists of essays written in 25 min by first-year undergrad students, without using external references. The essay assessment was done by different AWE tools such as Writing Assessment Tool, and Tool for the Automatic Assessment of Cohesion. The obtained results indicate that essays with many original ideas (flexible and elaborated) got a high evaluation and were significant features for determining the quality of essay. In our work, we evaluate elements of a conclusion, as those described in the pattern, with the aim to help students improve their writings. Similarly, as the work described previously, our research identified that the conclusions of graduate level obtained high values of connection to the objective, these being more extensive than those of undergrad level.

In a collected corpus of research proposals and theses, we found that conclusions that obtained high values (Goal Connectedness/Judgment/Speculation) after the evaluation corresponded to graduate students. These results suggest that graduate students with better writing skills (lexical richness) (González-López and López-López 2015) also achieved satisfactory results in the attributes examined in conclusions. Hence, the students who successfully completed a master or doctoral degree seem to possess better writing skills than students of college level. In addition, the result of a pilot test supported the conclusion that the students of the experimental group obtained better results than those in control group, when guided in the conclusions preparation.

Methodology and corpus

The first step of our study was the creation of a subcorpus of the Coltypi (http://coltypi.org/) collection which contains student theses, project and research reports. The collection includes documents of Graduate level: Master (MA) and Doctoral (PhD) degree; and Undergraduate level: Bachelor (BA) and Advanced College-level

Technician (TSU) (a two-year technical study program offered in some countries). The corpus domain is computing and information technologies. Each item of the collected corpus is a document (in Spanish) evaluated previously by a committee.

For each conclusion of the collection, the associated general objective was gathered. In total, 312 conclusions-objectives pairs (see Table 1) were obtained. Also, we can notice that on average, the conclusions of graduate level are longer than those of undergraduate level. However, the objective section tends to be shorter than conclusions section. To validate our model, 30 conclusions were selected with their corresponding objectives, 15 of bachelor and 15 of TSU level. Each conclusion was manually reviewed for the three elements by annotators.

The annotation process included two annotators, marking the text that reveals the presence of Goal Connectedness, Judgment and Speculation. Each of our annotators had experience in theses review. Table 2 includes an undergraduate objective-conclusion example tagged by the annotators, where S1 denotes Sentence 1. The annotators consider the objective (S1) as the pivot sentence, and then the annotators identify the connection between both sections.

Below, some sentences of undergraduate objective-conclusion pairs tagged by the annotators are provided.

Goal Connectedness (GC) text marked by annotators in a conclusion section:

S3: As we noted earlier, each driver manufacturer has a different method of accessing the internal information, therefore for this reason, the software designed should be adapted to the driver manufacturer, considering slight changes in the routing of the items (variables) located within the controller memory.

S4: The graphical interface designed is a clear example of the scope that has Visual Basic for design automation technologies and hence their wide use by international designers.

Speculative text marked by annotators in conclusion (SsP):

S6: Furthermore, as recommendation observe that the GUI can be modified at any time with the right software, with the use of the OPC library (open technology).

For Judgment Model the annotators only write: *Yes or Not presence of Judgment* The annotator task is complex since each academic reviewer has his own criteria for tagging, adding a certain level of subjectivity to the task.

The Kappa agreement between annotators for Goal Connectedness was 0.923 which corresponded to "almost perfect" (Landis and Koch 1977). For Speculation was 0.650 which corresponded to "substantial". Finally, for Judgment, the agreement was 0.72 (also "substantial").

Table 1 Text Corpus (words in average)

Level	Objective-Conclusion	Words in Conclusion	Words in Objective
Doctoral	26	584	37
Master	126	577	35
Bachelor	101	419	44
TSU	59	353	40

Table 2 Undergraduate objective-conclusion pair example

Objective:

S1: Design application software in Visual Basic for data acquisition of digital drivers using OPC technology.

Conclusion:

S2: This work shows the communication between software and PLC Allen-Bradley Compact Logix, covering processing needs for level control of a boiler.

S3: As we noted earlier, each driver manufacturer has a different method of accessing the internal information, therefore for this reason the software designed should be adapted to the driver manufacturer, considering slight changes in the routing of the items (variables) located within the controller memory.

S4: The graphical interface designed is a clear example of the scope that has Visual Basic for design automation technologies and therefore is widely used by international designers.

S5: Moreover, it can be noticed that the Ethernet communication provides a higher speed compared with the RS-232, using Ethernet we achieve a more reliable monitoring since we satisfied with the information presented on screen, achieving more efficient supervisory control.

S6: Furthermore, as recommendation observe that the GUI can be modified at any time with the right software, with the use of the OPC library (open technology).

S7: The interface turns into a tool that efficiently makes the control of a process within any industry to provide the operator updated and organized information, we mention that the basis of this program can be used for control of different variables either temperature, flow or pressure.

S8: Thus, we see that the OPC technology offers a variety of tools for client-server connection, showing great amplitude data management.

Model overview

The second step was the construction and model evaluation for the conclusion section. Our Model has a Conclusion Analyzer, which contains three main schemes (see Fig. 1) and seeks to help students with little or partial experience in drafting conclusions, to assess the elements that academic advisors deem important. In addition to the Conclusion Analyzer displayed on our model, we also include student feedback and recommendations. The suggestions are provided to the student, depending on the level reached in each of the attributes evaluated. Each of the recommendations was formulated by our annotators, which are higher education instructors with experience in research report and thesis review.

Goal connectedness scheme (GC)

This scheme seeks to identify whether the conclusion shows some connection with the general objective. The expected results are that some sentences display this relation,
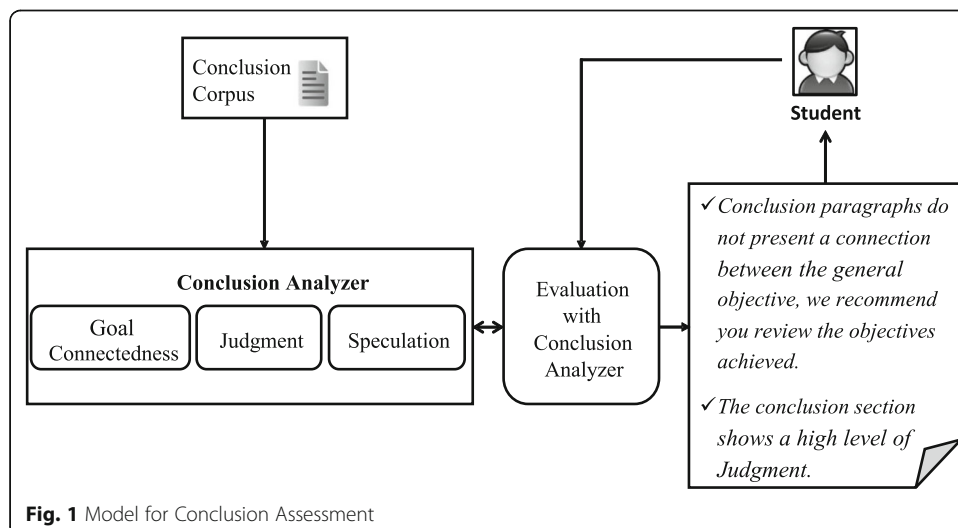

Fig. 1 Model for Conclusion Assessment

especially those at the beginning. So, the target such relations looking for the sentence that best cover the objective. In the first step, we remove function words in input documents, i.e., in conclusion section and general objective. Function words, also called stop words, include prepositions, conjunctions, articles, and pronouns. Also, each term was stemmed with FreeLing (nlp.lsi.upc.edu/freeling), a library of automatic multilingual processing functions, that provides analysis and linguistic text tagging. For the conclusion section, a group of sentences were employed, while in the objective section the full text was used, i.e. we consider an objective as one sentence. For computing the Connectedness attribute, this is done in terms of coverage, applying the expression in Table 3. To evaluate the GC, each of the objective-conclusion were processed in pairs with the Goal Connectedness scheme and the result was placed in a scale. To build the scale, the graduate texts were used as a reference, i.e., each objective-conclusion pair was processed, and after that, the average of all results was computed. However, to smooth out the scale, a group of 50 elements of undergrad level was included (selected at random).

Finally, to validate the scale, the corpus tagged by annotators was employed. After evaluation of the annotated corpus (30 objective-conclusions), the Fleiss Kappa between our analyzer and annotators was computed, obtaining a 0.799 value, corresponding to a "substantial" agreement. Below we present an example of the objective and conclusion of a thesis analyzed by the GC schema embedded in our tool. In these examples, the coincident words are underlined.

Objective

Design an *intelligent agent capable* of *interacting* with a person verbally and in writing in Spanish that helps the *vocational* guidance process.

Conclusion segment: ... as a finished product an *intelligent agent* that simulates a *vocational* counselor in an interview *capable* of *interacting* verbally and in writing in the Spanish language.

Judgment scheme (JS)

The goal of this scheme is to identify whether the conclusion section shows evidence of opinions, as the following conclusion: *It was demonstrated that the use of conceptual graphs and general semantic representations in text mining is feasible, especially beneficial for improving the descriptive level results.*

Table 3 Formulas and Parameters

Model	Parameters	Expression
Goal Connectedness	Absence of connection < 0.12 0.12 < Acceptable < 0.41 Strong connection > 0.41	$C = \frac{\#(So \cap SC_i)}{N}$ C = Coverage So = List of words in objective SC_i = Sentence i of conclusion N = Number of terms in the objective
Judgment	No Judgment < 7.84 Yes, presence of Judgment > 7.84 < 26.98	$T = \sum Wi \left(\frac{On+Op}{N}\right)$ T = Score[a] On = Negative Score; Op = Positive Score N = Number of occurrences (noun, pronoun) Wi = each word of sentence

[a]The result is obtained by adding the average load of each word of the sentence

To consider terms that reflect an opinion or value judgments, we turned to Senti-WordNet 3.0 since there is no such extensive resource for Spanish. The tool is a lexical resource for English, which aggregates an opinion score to each term (e.g. noun, adjective) depending of the sense. The sense has three numerical scores for objectivity, subjectivity and neutrality. The range of values is between 0 and 1. Each conclusion was translated to English employing Google Translator (A study of four services using Spanish to English translation showed that Google was superior (Aiken et al. 2009)). After translation, empty words were removed and the value for each sentence was computed. To obtain the measure of each sentence, each term was searched in SentiWord-Net. To evaluate the JS, we took again the graduate level texts as reference to define a scale. However, in this case, the smooth was not applied, as there are three levels of opinion. For this attribute, the conclusions must reach the average level of review, this will give evidence that the student is expressing judgments and opinions in the conclusion paragraphs. Fleiss Kappa was computed between the results of our analyzer and annotators (30 objective-conclusions pairs), reaching 0.65, a "substantial" level.

Speculation scheme (SpS)

The model aims to identify evidence of sentences that describe future work or derivations of the research.

For this purpose, two lists of speculative terms were employed. The first list has lexical features provided by (Kilicoglu and Bergler 2008), that includes modal auxiliaries, epistemic verbs, adjectives, adverbs, and nouns (see Table 4). The second list was obtained from the "Bioscope corpus", consisting of three parts, namely medical free texts (radiology reports), biological full papers, and biological scientific abstracts. Both lists are independent of our corpus. The dataset contains annotations at the token level for negative and speculative keywords (Vincze et al. 2008). The corpus was tagged by two independent linguists following guidelines. After extraction of speculative terms, the two lists were combined, with the goal of gathering a more exhaustive list. Each term of the merged list was translated, producing a list of 227 speculative terms.

Next, a conclusion segment is provided, including an example of speculation in the phrase "could be" that is underlined: *One of the applications in which this methodology could be used is the search for images using the image itself as a search parameter.*

To evaluate the Speculative attribute, each of the conclusions was processed counting the speculative terms in each sentence. Only the coincidence level between the text marked by the annotator and the sentence with maximum number of Speculation

Table 4 Speculative words

Feature type	Speculative words
Modal auxiliaries	may, might, could, would, should
Judgment verbs	suggest, indicate, speculate, believe, assume
Evidential verbs	appear, seem
Deductive verbs	infer, deduce
Adjectives	likely, probable, possible
Adverbs	probably, possibly, perhaps, generally
Nouns	possibility, suggestion

terms was described. After analyzing the annotated pairs using the criterion described, the Fleiss Kappa was computed between the results of our analyzer and the annotators (30 pairs), obtaining a result of 0.887, i.e. "almost perfect" agreement.

Academic level analysis

An exploration of the three selected attributes was done, as a way of validating the conclusion pattern. The whole corpus was explored, identifying the position of the attributes Goal Connectedness, Judgment and Speculation, for the different academic levels. According to the Conclusions pattern, the Connectedness is located at the beginning, the Judgment at the center and future work (Speculation) at the end of the conclusion.

The percentages found for Goal Connectedness-Judgment and Judgment-Speculation are presented (see Table 5). The percentage (Found) represents the number of conclusions where comparisons were done, i.e. the similarity between the conclusions and the conclusion pattern was identified, otherwise included as (Not found). In Table 5, we note that the graduate level has a higher percentage than undergraduate level, i.e., postgraduate students wrote the conclusion section adhering to a structure. The structure tends to relax in undergraduate (BA and TSU) writings.

The Pearson correlation coefficient between Goal Connectedness and Judgment was of 0.65. The correlation value between Goal Connectedness and Speculation was 0.17. Between Judgment and Speculation, the correlation was 0.28. This level of positive correlation suggests that the presence of Goal Connectedness, and Judgment attributes is common in the conclusions.

The higher connection level of the objective with the conclusions section increases the level of judgments or opinions. One can infer that the two elements are relevant to the student when writing the conclusion. However, the Speculation attribute will not necessarily appear in the conclusions as some students write future work in separate sections.

Conclusion analysis in practice

After the corpus exploration and evaluation of methods to assess conclusions, an online system was developed with the goal of validating the models and identifying if the tool could help students to improve their writings. The computational tool [anonymized n.d.] (In Spanish: Tutor Revisor de Tesis) is hosted at tutor.turet.com.mx. Any student can register and use the system. In addition, [anonymized n.d.] has a section that explains its use and provides support material for the student. The support material gives the student an explanation of the elements evaluated by the system.

Table 5 Explored Attributes in whole corpus

Academic Levels	Goal Connectedness - Judgment	Judgment - Speculation	Result
PhD	88%	76%	Found
	12%	24%	Not Found
Master	79%	71%	Found
	21%	29%	Not Found
Bachelor	64%	63%	Found
	36%	37%	Not Found
TSU	58%	67%	Found
	42%	33%	Not Found

Figure 2 shows the main interface of the system where the student submits the objective and conclusion of his/her report. Subsequently, the system sends the results of the analysis back to the student indicating if the score reached is acceptable. The student can repeat the analysis and each attempt is recorded. For instance, in case of no evidence of Judgment, the system provides the following text "Opinion is very important in a conclusion, to achieve an acceptable level of judgment, improve the conclusion by incorporating sentences that contain your value judgments". In case of Goal Connectedness was strong, the system sends the message "The connection value is strong between your objective and your conclusion. Congratulations, you have achieved an excellent score". The system was created with Django, Python, and libraries for text analysis.

Pilot test

We designed and performed a pilot test to assess the impact/benefit of using an online application focused on Goal Connectedness, Judgment and Speculation in a conclusion section of a research report. The experiment involved undergraduate engineering students. Also, two randomly selected groups were considered, one experimental, and other for control, each with 15 students. The two groups received instructions on how to write a conclusion section. Students were informed of each essential attribute, using the triangle pattern of conclusion section. The control group had a traditional monitor, that is, an academic advisor reviewing their documents, while the experimental group had access to the intelligent tutor 24 h a day.

All documents produced by both groups were evaluated with [anonymized n.d.] to compare the results among them. The foremost hypothesis to be validated in this pilot test was: "The use of an online application, allow students in the experimental group generate documents with better parameters, in terms of Goal Connectedness, Judgment and Speculation". Table 6 depicts the measures obtained in the pilot test for the two groups.

One can notice that the experimental group produced higher values on each attribute than control group. These results provide evidence that students of experimental group reach twice the values of measures. It was also observed in the experimental group that on average, the number of attempts of [anonymized n.d.] use was 8. However, when
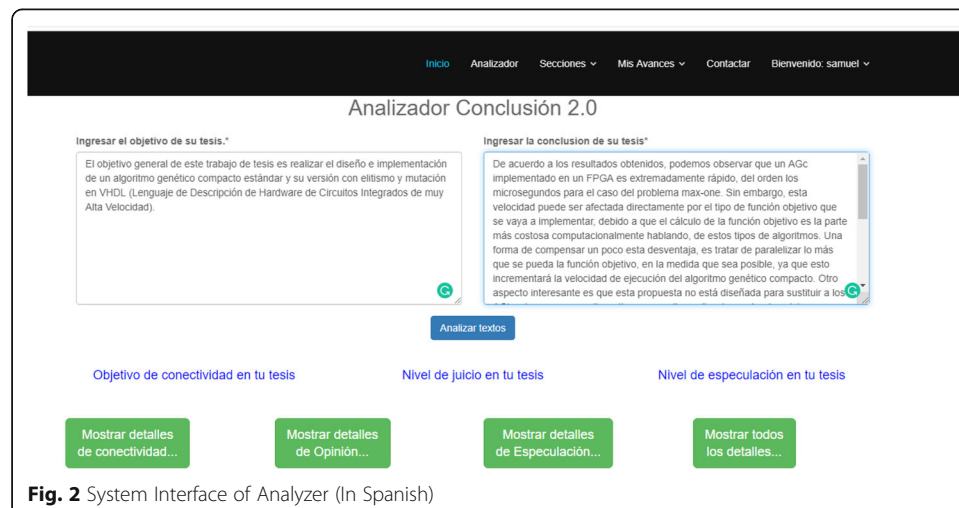

Fig. 2 System Interface of Analyzer (In Spanish)

Table 6 Measures obtained by both groups

Groups	Goal Connectedness	Judgment	Speculation
Experimental	4.07	81.08	25.00
Standard Deviation	0.22	5.19	1.40
Control	2.45	39.20	14.00
Standard Deviation	0.14	1.29	1.15

the standard deviation was observed in the control group, we found that it was lower than the experimental group. This could indicate that the control group is more uniform in performance. It is possible that in the experimental group some students using a technological tool allow them to achieve superior results, while other students have an average performance on the test.

Also, a statistical analysis to validate the results was performed. We applied a hypothesis test for two independent samples with different standard deviation. The confidence level was 95%. The hypothesis test for each measure was carried out. For the three attributes, the null hypothesis was rejected with p-values of 0.046 (Goal Connectedness), 0.020 (Judgment), and 0.024 for Speculation attribute. These statistical results indicate that the null hypothesis is rejected for the three characteristics. The [anonymized n.d.] system allowed students to achieve higher measures than the students in the control group.

In addition, a correlation analysis was performed among the three characteristics in the two groups. The aim of this analysis is to identify the level with which the three analyzed elements are close, according to the pattern of conclusion described above. In addition, the results between both groups were compared, with the expectation of finding higher correlation in students of the experimental group derived from the use of the computational resource.

In Table 7, one can observe a correlation of the experimental group which is quite close to the correlation identified in the annotated corpus. The characteristics of Goal Connectedness-Judgment show a positive correlation with significance in the annotated corpus and in the experimental group, i.e. a value of 0.609. The result of Goal Connectedness-Speculation shows that there is no correlation, as is the case of the annotated corpus. We can assert that the students wrote conclusions with a closeness to the pattern of conclusions, since the correlation numbers were close to those of the annotated corpus.

For the students of the control group no correlations were found, which indicates that control students should continue working with the writing of their conclusions, to reach acceptable values.

A satisfaction survey based on Technology Acceptance Model (Tobing et al. 2008) was also applied, to assess the opinion of the experimental group on using the online analyzer, in the aspects of usefulness, ease of use, adaptability and intention to use the system. Students answers were based on a five-point Likert scale ranging from 1 ("Strongly disagree") to 5 ("Strongly agree"). We can observe in Fig. 3 results in terms of averages by aspect of the satisfaction survey, where the preference of the students is above 4 points ("Agree") for each aspect, so one can conclude that the analyzer was found useful, easy to use, adapted to their level and students have the intention to use it. However, in student comments it was found that some felt the registration was complex, primarily as it had a traditional registry process which requested confirmation via email.

**Table 7** Experimental and Control Group Correlations

Attributes	Experimental Group		Control Group	
	Correlation	*P*-value	Correlation	*P*-value
G. Connectedness-Judgment	0.609	0.016	0.36	0.187
Judgment-Speculation	0.535	0.04	0.042	0.881
G. Connectedness-Speculation	0.223	0.424	0.339	0.216

Also, in Fig. 3, it can be observed that the tool was useful to students; however, the intention to use, despite being above "4", this can be considered as a weak aspect that can be improved. We can conclude that the analyzer was found useful, easy to use, adapted to their level and students have the intention to use it.

Conclusions

A system has been presented that uses natural language processing techniques. The system is designed to consider specific attributes of writing in a conclusion section suggested by authors of methodology books and institutional guides. In our work, we take advantage of the knowledge in the theses in our corpus, previously reviewed by different academic advisors, when extracting the attributes with distinct proposed models. It was found in the annotated corpus that postgraduate level student texts outperformed those of undergraduate level across the three essential attributes. The behavior provides evidence that students with more practice writing a research reports or thesis (graduate level), possess better skills. Furthermore, our models can help improve the writing of research report by undergraduate students or inexperienced learners, in relation to the attributes of Goal Connectedness and Speculation, since the achieved Kappa levels were substantial or better.

The pilot test with engineering students in the systems area allowed us to bring the developed models to a real environment. We can identify, as a result of the pilot test, that

Fig. 3 Scores of Control and Experimental Groups
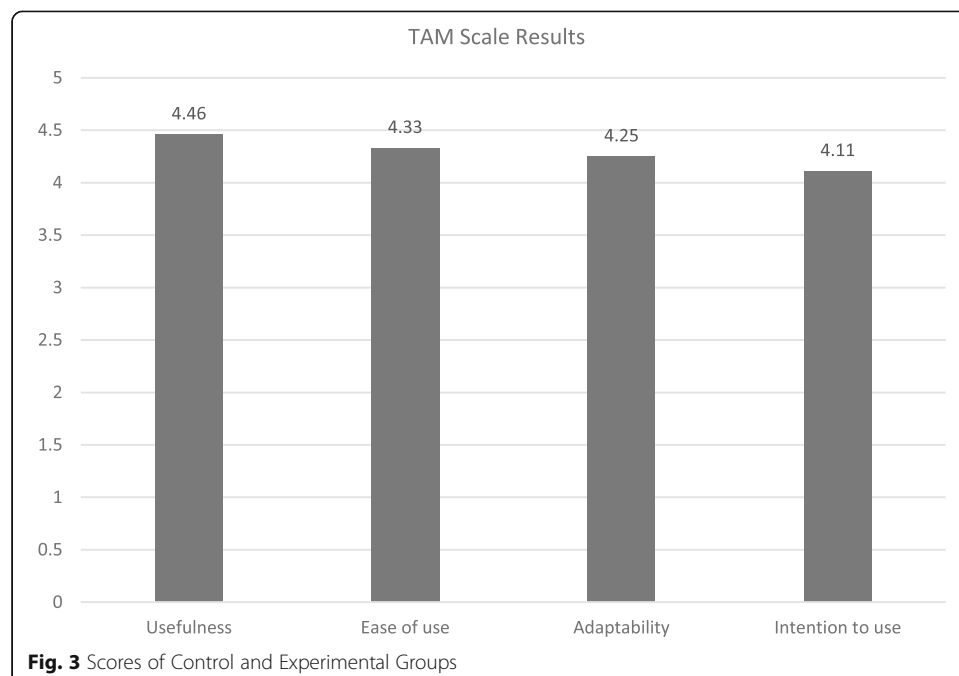
the students of the experimental group showed interest in using the tool and improving their writing. Such interest was observed in the average number of the times that the students used [anonymized n.d.]. However, it could have also been due to the competition generated amongst the students of the experimental group when using the system, as results can be improved when using the tool. A special case was a student who used [anonymized n.d.] but with a text very distant from the project he was doing, perhaps only to comply with the use of the tool. In the short term, the tool will be improved regarding its registration process, allowing the use of social networks as access to the system.

One of the constructs that were best evaluated in the satisfaction survey was the usefulness that motivates us to continue with this project. The intention to use construct was the lowest, so strategies to increase this metric were sought, for example, the incorporation of serious games (Long and Aleven 2014). We also plan to incorporate a section where the students can check their progress graphically.

The results of the correlation analysis between the two groups (control and experimental) validated to some extent the similarity with the pattern of conclusions detailed in the introduction. One finding was that the Goal Connectedness and Judgment measures showed a positive correlation with significance, such as that found in the annotated corpus, where the documents were theses or research projects reviewed previously by a qualified committee.

Furthermore, there are also plans to include metrics to assess whether a conclusion contains a certain level of originality and elaboration. The working hypothesis is that the conclusions of graduate level contain more original ideas than undergraduate level. For future work, we plan to extend the analysis to consider speculative phrases and include in our reference corpus examples of thesis of social sciences. With the results obtained in this research, the system [anonymized n.d.] can be a tool that precedes the task of the academic reviewer and helps the student his or her drafting of research reports, theses or scientific documents. In addition, a deep analysis will be performed to identify if the feedback provided by our model has a positive impact on the learner.

References
M. Aiken, K. Ghosh, J. Wee, M. Vanjani, An evaluation of the accuracy of online translation systems. Commun. IIMA **9**(4), 67–84 (2009)

G.R. Allen, *The Graduate Students' Guide to Theses and Dissertations: A Practical Manual for Writing and Research* (San Francisco, Jossey-Bass Inc., 1973)

S. González-López, A. López-López, Lexical analysis of student research drafts in computing. Computer Applications in Engineering Educcation 23(4), 638–644 (2015)

J. Bitchener, H. Basturkmen, Perceptions of the diculties of postgraduate l2 thesis students writing the discussion section. J. Engl. Acad. Purp. **5**(1), 4–18 (2006)

S.A. Crossley, K. Muldner, D.S. McNamara, Idea generation in student writing: Computational assessments and links to successful writing. Writ. Commun. **33**(3), 328–354 (2016)

S.A. Crossley, L.K. Varner, R.D. Roscoe, D.S. McNamara, in *Conference on Artificial Intelligence in Education*. Using automated indices of cohesion to evaluate an intelligent tutoring system and an automated writing evaluation system (Berlin, Springer, 2013), pp. 269–278

J. Davis, R. Liss, *Effective Academic Writing 3* (New York, Oxford University Press, 2006)

J.C. Debuse, M. Lawley, R. Shibl, Educators' perceptions of automated feedback systems. Australas. J. Educ. Technol. **24**(4), 374–386 (2008)

M.J. Gierl, S. Lati, H. Lai, A.P. Boulais, A. De Champlain, Automated essay scoring and the future of educational assessment in medical education. Med. Educ. **48**(10), 950–962 (2014)

H. Kilicoglu, S. Bergler, Recognizing speculative language in biomedical research articles: A linguistically motivated perspective. BMC bioinformatics **9 Suppl 11**(Suppl 11), S10 (2008). https://doi.org/10.1186/1471-2105-9-S11-S10

J.R. Landis, G.G. Koch, The measurement of observer agreement for categorical data. Biometric **32**(1), 159–174 (1977)

Y. Long, V. Aleven, in *Conference on Intelligent Tutoring Systems*. Gamification of joint student/system control over problem selection in a linear equation tutor (Cham, Springer, 2014), pp. 378–387

D.S. McNamara, S.A. Crossley, P.M. McCarthy, Linguistic features of writing quality. Writ. Commun. **27**(1), 57–86 (2010)

P.O.W Lab: Purdue Online Writing Lab, *Introductions, Body Paragraphs, and Conclusions for an Argument Paper* (2018) Resource document. https://owl.purdue.edu/owl/general_writing/common_writing_assignments/argument_papers/conclusions.html. Accessed January 30

V. Tobing, M. Hamzah, S. Sura, H. Amin, Assessing the acceptability of adaptive e-learning system. Int. J. Comput., Internet Manag **13**, 3 (2008)

V. Vincze, G. Szarvas, R. Farkas, G. Móra, J. Csirik, The BioScope corpus: Biomedical texts annotated for uncertainty, negation and their scopes. BMC bioinformatics **9 Suppl 11**(Suppl 11), S9 (2008). https://doi.org/10.1186/1471-2105-9-S11-S9

Publisher's Note