Supporting the Review of Student Proposal Drafts in Information Technologies

Samuel González López and Aurelio López-López
Instituto Nacional de Astrofísica, Óptica y Electrónica
Luis Enrique Erro No. 1, Tonantzintla, Puebla México
52 + (222) 2663100 Ext: 8314
{sgonzalez,allopez}@inaoep.mx

ABSTRACT
In many cases, academic programs or courses conclude with a thesis or research proposal text, elaborated by students. The review of such texts is a heavy load, especially at initial stages of drafting. This paper proposes a model that allows linguistic and structural review of some essential elements in proposal drafts of undergraduate students. The model aims to support the review from vocabulary to the argumentation in the draft, and is part of an intelligent tutor to monitor student progress. This work presents the initial results in terms of lexical and global coherence analysis of proposal drafts of students. Lexical analysis is done in terms of lexical density, lexical diversity, and sophistication. Global coherence is evaluated using the Latent Semantic Analysis technique. Our results show that the level reached so far by the analyzer is adequate to support the review, taking into account for one section the level of agreement with human reviewers.

Categories and Subject Descriptors
K.3.2 [Computers and Education]: Computer and Information Science Education – Computer Science Education.

General Terms
Measurement, Experimentation, Languages, Verification.

Keywords
Student drafts, lexical density, lexical variety, sophistication, coherence, evaluation, intelligent tutor, language models.

1. INTRODUCTION
Academic programs or courses conclude often with a thesis or research proposal text, elaborated by students. The usual process that students follow is to write a first draft and then improve the document with the iterated recommendations of the adviser. Some educational institutions have a guide that supports students in structuring the proposal document, however in many cases this is insufficient. It was observed that students often need help on how to structure all aspects of their draft. This requires that the academic adviser or instructor spends extra time in the review process. This paper focuses on developing a computational model that helps undergraduate students of computing and information technologies area, to improve their draft during the writing process, especially in the early stages. Also we intend that this model implemented in a system helps the academic adviser by reducing the time dedicated to the draft review, focusing on the content. The results reported here on lexical analysis and global coherence, are part of a larger project that may help students to evaluate early their drafts, and facilitate the review process of the academic advisor.

The proposed model consists of four levels: where the first focuses on the lexicon used by the student in his/her draft, the second level seeks to identify and assess the level of coherence, the third level considers language models intended to identify the particular structure of each element of the proposal, and the last one focuses on identifying answers to methodological questions that characterize certain elements of a proposal document, for example: What will you do?, How are you going to do it? These levels of evaluation will be managed by an intelligent tutor that will provide feedback to the student. We selected eight essentials elements of a research proposal: title, problem statement, justification, objective, research questions, hypothesis, methodology, and conclusions.

We report in this paper some initial results about evaluating global coherence and the lexical analysis in proposal drafts. Our initial results show that our analyzer provides adequate review, considering the level of agreement with human reviewers.

2. LEXICAL AND COHERENCE ANALYSIS
This section reviews the main work in the area of natural language processing (NLP) techniques related to lexical analysis including some measures employed, and for coherence assessment.

2.1 Lexical analysis
There are a variety of methods to evaluate the use of vocabulary (lexicon) in text, all with different goals. To measure the sophistication of some papers using a text word lists, in [1], they used a list of 3000 easy words. For Spanish, some studies use the list provided by the SRA\(^1\) (Spanish Royal Academy) of 1000, 5000 and 15000 most frequent words. In [2], 32 lexical measures were used to predict demographic attributes, such as age or gender regardless of the domain. Those measures were grouped in lexical diversity, lexical density, and sophistication.

Others works have used Yule’s K to measure the richness in texts [3], where this kind of measures focus on the word repetitions and

---

\(^1\) Most frequent word list available at http://corpus.rae.es/
is considered a measure of lexical diversity. In our work this assessment is on the basis of the evaluation process of a proposal draft. We are looking for lexical analysis to identify frequent deficiencies in student writings such as too much use of empty words, abuse of certain terms, or low knowledge of technical terminology. Once identified these common errors, the tutor can provide feedback.

### 2.2 Coherence analysis

A formal definition given in [4] establishes that coherence is the connection of all parts of a text into a whole: the interrelationship of the various elements of the text. Also coherence is classified based on its scope: local and global. The global coherence means that a document is related to a main topic, i.e. it is not consistent when its elements have no such main topic. And local coherence is defined within small textual units [5]. Therefore, coherence within proposal draft is important because if it does not have each of the elements related into a whole or sections are not close to a topic, the document would not exhibit coherence. Different techniques such as Latent Semantic Analysis (LSA) and Grid Entity have been used to assess coherence. The first focuses on the semantic aspect and the second in syntactic features.

In [6], their work assessed automatically the coherence of police news, that is, given police news reports written by a columnist, the evaluation system provided the degree of coherence in the reports. This research used the technique of Latent Semantic Analysis, first compiling a police news corpus, which served to train the system. From this set of texts, the system evaluator measured the coherence of the news. The expected result was that the software evaluator will approach the results of evaluations obtained by a columnist and a Spanish language expert.

In our work, we also applied the LSA technique to assess the global coherence of the sections in proposal drafts, to provide a measure of coherence to the student about his/her writing.

A representation of discourse called Entity Grid is presented in [7], which is constructed in a two-dimensional array that captures the distribution of entities in the discourse across sentences, where rows correspond to the sentences and the columns represent the entities of discourse. The cells can have values, such as: subject, object or neither. The main idea of this representation is that, while the object and subject are present in the sentences, the assessed coherence is stronger. They assume that certain types of transition of subject and object are likely to appear in locally coherent discourse. The Entity Grid technique is intended to reveal local coherence and compared to LSA in terms of correlation [8], both techniques showed low correlation, so they capture different aspects of coherence.

A combination of algorithm BL08 (nouns and pronouns) for entity grid with writing quality features, such as grammar, word usage, and mechanics errors, showed improvements in the review of the coherence of student’s essays on three different populations [9]. The experiments used a corpus of 800 essays related to Test of English as a Foreign Language (TOEFL) and the Graduate Record Admissions Test (GRE). After performing the experiments, only two out of three populations obtained acceptable Kappa values, between humans and system.

### 3. INTELLIGENT TUTORS

A system called AutoTutor is presented in [10], an intelligent tutorial, which simulates the pattern of discourse and the teaching strategies of a human tutor. AutoTutor modules have a set of processors and storage units that keep the content dynamic with qualitative and quantitative parameters. The aim of AutoTutor is to build an intelligent agent that can produce effective conversational dialogues pedagogically, using natural language conversations to help students actively to construct knowledge.

The tutor is designed to teach Newtonian physics to students, and starts with an introduction and an overview of the subject of two minutes, and then moves to the learning session. After the introduction, the conversation begins by establishing a problem and asking the student for an answer. From this point, it starts a conversation between the student and the tutor. The goal is that students using the intelligent tutor can answer the questions correctly. The outputs of AutoTutor are stored and updated during the process of mentoring, in this way it creates and updates the tutor for each student, by the history of their dialogue. This tool uses the Latent Semantic Analysis (LSA) technique to represent semantic knowledge of physics, and allows a comparison of the contributions of the student to the expected responses. This type of work focuses on creating an environment similar or close to what the student would have with a teacher who provides help.

One of the artificial intelligence techniques to model the student in intelligent tutoring is Bayesian networks. Similar to our proposal, the paper [11] presents the modeling of the students, integrating the Bayesian networks and the tools of the relational database model. In modeling the student, the cognitive state is inferred from two parts: the data prior to the student and the behavior during the interaction with the system.

A tutor that allowed the student to solve thermodynamic problems is reported in [12]. The student can select a problem and solve it in the workspace. Solving such problems involves two main phases: drawing the diagram, and later calculating unknowns, the solution can be sent to the tutor for evaluation and feedback. This selection feature will be incorporated into our intelligent tutor for the student to evaluate their partial or global proposal and receive feedback.

### 4. ANALYZER MODEL

Our model incorporates a module for linguistic and structural review, both will be integrated into an intelligent tutor, which allows students to get feedback and adapt the information displayed depending on their performance.

Currently, we have a corpus of 50 thesis proposal documents stored in database, these research projects are in Spanish and mainly from the Computing and Information Technologies domain. The topics included in the corpus are diverse, some documents are in the area of computer architecture, intelligent tutoring, machine learning, information retrieval, communication networks, and software engineering. The coherence and lexical analyzer is embedded in an intelligent tutor using Bayesian network technology to build the student model.

The intelligent tutor generates a model for each student and the model is updated each time the student uses the analyzer, the aim is that the tutor can monitor student performance and send feedback about their proposal draft. The analyzer model consists of the following modules.

*Reached level identifier:* This identifies the level of student when he/she enters the intelligent tutor, accessing a repository of previous levels achieved by the student to extract the most recent.
This level lets the tutor to establish a scenario according to the student’s progress, for instance suggestions, material or recommendations that the tutor displayed while the student uses the analyzer. When the user utilizes the tutor for the first time, he/she starts with a low level and a basic scenario with definitions, for example, of concepts of coherence, lexical diversity, lexical density, or sophistication. In Figure 1, we illustrate this module.

![Analyzer Model](image)

**Figure 1. Analyzer Model**

**Achieved levels:** It is a repository for saving the student reached level and updating each time the student makes an assessment of the elements of his/her proposal draft.

**Analyzer:** This module contains two types of reviewers. The first, the Linguistic reviewer, has the purpose of analyzing the different elements considered as essentials in a proposal draft. The aim is to evaluate linguistic characteristics of each element. The second, the Structural reviewer, seeks the coherence of the different elements and the relation that there has to exist between them. The aim is to give a result of a global analysis of the elements of a research project proposal document.

**Results and errors analyzer:** This module focuses on adapting the tutor to the student, i.e., in taking into account the student actions and the results produced by the reviewers to generate the level of achievement. For this module, the frequencies of mistakes are considered.

**Feedback generator:** This component will provide different recommendations to the student, based on the performance achieved. The goal is that the intelligent tutor can supply positive or negative example to illustrate something wrong, this decision will depend on the results that the student is generating using the tutor. For instance, when a student receives a low evaluation of overall coherence in the objectives section of the proposal, the tutor sends two examples and these let students identify how an appropriate objective has to be written, and also sends a negative example for comparing them and see the differences. For instance, a positive example of global coherence2 “Explore the requirements for sharing information through a search and retrieval system among interoperable digital libraries with the open archives initiative, using mobile agent”. A negative example of global coherence “Inquire what is the role of school cooperatives in the government of national schools, analyzing the scope and limits of family involvement in schools and the educational system”.

It also displays a brief explanation of both examples, negative and positive, with the aim of distinguishing the errors and good decisions of each.

**Case repository:** This contains positive and negative examples of each section in proposal drafts, grouped by each aspect that is evaluated. The examples were taken from thesis of research institutes and universities with computing and information technologies programs. The repository will be updated with the same elements evaluated, once they obtain high measures and the approval of academic adviser.

**Linguistic and Structural reviewers** will be based on a four levels assessment. This module starts with a basic assessment and goes to a more complex level.

![Four-level evaluation](image)

**Figure 2: Four-level evaluation**

The eight elements of project proposals to evaluate are illustrated at the bottom of Figure 2, the first level is a lexical analysis. For a full assessment at this level, there are three dimensions, the first is the lexical diversity which seeks to measure student ability to write their ideas with a varied vocabulary, the second dimension refers to the lexical density whose goal is to reflect the proportion of content words after removing empty words, finally the sophistication reveals the knowledge of the technical subject and is the proportion of “advanced” or “sophisticated” words employed. Together, the three dimensions will identify the lexical level of the student writing. The sophistication would be a plus for the undergraduate student. In section 4, we describe the initial experiments and preliminary results about lexical analysis.

At the second level, an evaluation of coherence for the essentials elements of the document written by the student is done. This level incorporates the LSA and Entity Grid techniques to capture the global and local coherence. Currently, we have developed a

---

2 Examples are translations of actual Spanish texts in our corpus
coherence analyzer with LSA technique only, this analyzer performs an assessment of global coherence of each section horizontally, i.e. where each element is compared to a specific corpus for that element. We want to know if the item is coherent with respect to a latent semantic space previously generated as reference. In section 4, we describe the initial experiments and preliminary results for global coherence.

After the second level, we consider the creation of language models for each element of the proposal drafts. The language models are intended to consider a syntactic view, whose aim is to identify syntactic patterns independent of content but proper to the specific element (section). For instance, a general objective most often begins with a verb in infinitive, subsequently displaying a noun with some additional clause. This type of characteristics would be revealed by the language models. In summary, the idea is to capture the particular syntax used in each element of proposal drafts.

In the fourth level, we propose to develop a method to identify the existence of answers to methodological questions. This question receives as answer a word sequence that makes sense to what is asked, for example, a methodological question to formulate the objective is What will you do?. The answer requires the use of a verb expressing an action and an object of the verb that indicates the context on which execute the action. This context is diverse and involves a variety of words to use. Using the following example, we illustrate the answers to methodological questions:

Develop an inductive learning algorithm for solving binary classification problems from unbalanced data sets, where the results obtained take appropriate reach consensus between accuracy and comprehensibility.

The question What will you do? would have the answer: “Develop an inductive learning algorithm”. The methodological question How are you going to do it? would require an appropriate answer specifying the way to achieve the goal. It can be observed that the answers consist of a sequence of terms. Identifying this type of answers is a challenge, since current studies have focused on very specific answers such as dates, places, people, or definitions [13]. Methodological questions are characteristic of elements such as objective and justification.

Also in the fourth level we have identified the argumentation as a feature that in the future could be assessed. The argumentation in a project proposal is relevant since it allows determining the resources used to argue on a position or idea expressed in the proposal draft by the student. Some works have studied how knowledge can build collaborative argumentation in online discussions [14]. Once achieved the four levels of evaluation, the intelligent tutor can send the student a score or a set of scores along suggestions of ways to improve the draft.

5. INITIAL RESULTS

We gathered a corpus of the different elements in proposal documents in Spanish. We distinguished in this corpus two kinds of student texts: graduate proposal documents, and undergraduate drafts. The first kind of texts includes documents already reviewed and approved by faculty, so they are considered as reference or training examples. The second kind of documents are used as test examples. The whole corpus consists of a total of 380 collected training samples and 80 test samples, as detailed in Table 1. The corpus domain is computing and information technologies.

| Table 1. Training and Test Corpus |
|-----------------|----------------|----------------|
| **Section**     | **Training**  | **Test**      |
| Problem statement | 40            | 10            |
| Justification   | 40            | 10            |
| Research Questions | 40            | 10            |
| Hypothesis     | 40            | 10            |
| Objectives     | 60            | 20            |
| Methodology    | 40            | 10            |
| Conclusions    | 40            | 10            |

5.1 Lexical Analysis

We are already evaluating student drafts at lexical level, i.e. the lexical analysis at the bottom of our 4-level evaluation. This analysis was performed along the three dimensions detailed in previous section, each in the range of 0 (worst) to 1 (best). Lexical diversity measures the number of types or different words used in the text. Density is computed as the proportion of lexical items and the total number of words in the text. Sophistication is computed as the percentage of word out of a list of common words (in our case in the 1000 common words according to SRA). Using the collected corpus, examples of undergraduate level were assessed in the three dimensions and compared against the values obtained for the graduate corpus. Figure 3 depicts the average of the three measures for the different elements, obtained for both subsets of the corpus. As expected, one can notice that graduate documents produced higher averages than undergrad drafts for the different elements.

In consequence, we obtained different ranges that define the scale for the eight sections of a draft. Applying the scale, we obtained the following examples of objectives with High and Low marks in the lexical analysis:

Objective (High level): Implement an algorithm based on hierarchical structures with volume enveloping of spheres for collision detection.

Objective (Low level): Create an information management system for franchises with relevant data of each establishment and personnel data of each franchise, as well as references of franchisees and personal of trust that manage the franchises.
One can notice that the first example is succinct and concrete, whereas the second example is quite verbose, with scarce technical terms, and abusing of the term franchise, lowering its lexical diversity.

5.2 Coherence

To assess global coherence using Latent Semantic Analysis in drafts, we set an experiment to validate our process. First, we asked three instructors to evaluate our whole collection of objectives (training and test subsets), eighty in total, see Table 1. We evaluate the level of agreement among evaluators. Then, we computed the semantic spaces for the different sections of our training subset and then evaluating automatically the objectives in the test subset. Finally, we evaluate the level of agreement between the grade assigned by the system and by instructors.

All the collection of objectives was sent for evaluation to three instructors serving as reviewers, that have experience in advising students in the preparation of their drafts in computing and information technologies. The reviewers did not know beforehand the level (graduate or undergraduate) of each sample. Each reviewer was requested to assign a level to each sample, using the scale: High, Medium and Low coherence, where the highest level meant that the text has a strong coherence or relationship to the domain of computing, and the low level means that the relationship is weak relative to the domain. An example of high coherence is: Analyze problems that arise in the system development of software architectures of Enterprise type. We can observe that the word “systems”, and “software” are very close to the domain, including the term “architecture” surrounded by the above terms fit within the domain of computing. Likewise, words with less thematic load such as “development” or “analyze” are close to the domain. We find low coherence in this example: Identify the effect of feedback on the learning of the business leader, to allow to be more effective. Notice that even though terms like “learning” or “feedback” may have some proximity to the domain, the words or phrases “business”, “leader” or “be more effective”, are the central topic and do not match the domain of interest.

The assessments provided by our reviewers allowed to exclude those examples in our training set considered low by at least two, or those where they did not agree, since they will bias the construction of the semantic spaces. On the other hand, the assessments on the test set allow comparing the automatic evaluation of coherence.

The Fleiss Kappa coefficient of agreement was computed for the three reviewers considering the test corpus. Table 2 shows the Fleiss Kappa results for each level, for the objective section.

<table>
<thead>
<tr>
<th>Kappa</th>
<th>Fleiss Reviewers</th>
<th>Cohen Coherence analyzer</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>0.6862</td>
<td>0.0000</td>
</tr>
<tr>
<td>Medium</td>
<td>-0.0378</td>
<td>0.2609</td>
</tr>
<tr>
<td>Low</td>
<td>0.7353</td>
<td>0.4218</td>
</tr>
<tr>
<td>Overall</td>
<td>0.5458</td>
<td>0.2237</td>
</tr>
</tbody>
</table>

3 Kappa(Landis y Koch, 1977)

The reviewers had a Substantial agreement for the Low and High grading, and a Poor agreement in the Medium. For the results obtained, we conclude that reviewers clearly identified High and Low levels. The overall level achieved between evaluators was 0.54, this giving Moderate confidence of agreement for the experiment.

We built the latent semantic space using the LSA technique with training samples of the whole collection, and in particular of objectives. The texts were processed by removing stop words and applying lemmatization using Freeling4.

Similarly as for the lexical analysis, we obtained the scale for automatically assessing coherence in each section, but here we used cross-validation, obtaining the levels High, Medium and Low, according to ranges of values produced by LSA. These levels allow to automate the evaluation of the coherence analyzer. In particular, for the objective section, we got an average of 0.49 with a standard deviation of 0.17, resulting in the highest threshold of 0.64 and the lowest threshold at 0.28.

Once the scale is defined, we evaluated the test samples with the aim to compare the results produced by human evaluators. In this case, Cohen’s Kappa is pertinent to compare the level of agreement between human and our coherence analyzer results. Table 2 shows the Cohen’s Kappa results for the human versus coherence analyzer.

We observed that the levels of agreement in the Low case is Moderate and Medium level is Fair, the overall level of agreement between humans and the analyzer was Fair. We conclude that the analyzer would have an acceptable support for the student and academic advisor in the process of preparing the proposal draft.

After comparing the statistical results, in terms of the Kappa coefficient of agreement, we also performed a qualitative analysis between the results of coherence analyzer and the process of reviewing a proposal draft, i.e. the advisor would expect that the analyzer was a first filter so that when the drafts reach him, at least have a Medium or High Level. Under this premise, the results of our analyzer match the concept of a strict filtering reviewer, because it provided low and medium values in most test objectives.

We can observe that if our system does not achieve at this time a higher level of agreement in the high level, this is not a problem since the analyzer is being stricter to assign the high level. In the experiment, the analyzer evaluated as Medium the few highest levels assigned by the reviewers. If the analyzer behaves more flexible and allows high level to objectives that have to be of a medium or low level, this could cause a burden to the academic advisor, failing to support in review.

Finally we note that between the coherence analyzer and human evaluators, the agreement is Moderate for low levels, bringing confidence that the analyzer is identifying those objectives that were classified as Low level for evaluators.

After assessing coherence, the analyzer sends feedback to the student for the seven selected sections in the draft, and updates the parameters of performance in the intelligent tutor. The analyzer can send recommendations, managed by the intelligent tutor. For example if the analyzer identifies a low level, the recommendation

4 This software is available at http://nlp.lsi.upc.edu/.
could be “It's suggested to restructure your objectives, making explicit the technical details of the computing and information technologies domain.”

These recommendations are handled by the intelligent tutor, being part of the student profile, with the aim to improve student writing.

6. CONCLUSIONS

In this paper, we have presented a model that combines natural language techniques and intelligent tutors. The model considers important features of writing at different levels.

In terms of vocabulary usage, we already can provide feedback when the draft is poor in terms of content words (i.e. excessive use of empty words), use of certain terms (many repetitions), or knowledge of technical terms. Also, we validate that graduate student showed a higher level in writing than undergraduate, in our corpus, regarding these three dimensions.

The LSA technique allowed evaluating the global coherence of objective section in proposal drafts, reaching an acceptable result of the percentage of agreement respect to human reviewers. It was crucial to have a gold standard to compare our results.

We will continue increasing the size of the corpus, so that the analyzer has a higher coverage, since the computing and information technologies domain is quiet extensive and growing.

In these initial experiments, the evaluation of coherence and lexical analysis was important to identify the student level, but could be improved by using all the four levels proposed for the evaluation. This will help students to improve their writing, and academic adviser would have more time to review the contents of the proposal documents. This model assumes that students have prior knowledge to write general documents, so it is not our interest to consider a grammar reviewer.

We expect that this computational tool generates in students a motivation to develop their proposal drafts and this analyzer will contribute to the advance in their proposal drafts. We currently have a web interface for the student to evaluate the draft in the first two levels, coherence and lexical analysis.

Bringing our model to a different domain does not seem too challenging, neither moving it to a different language, assuming similar language processing resources and corpus are available.

Currently, we are working on integrating the Entity Grid technique to evaluate the local coherence. We are also exploring the language models based on n-grams for characterizing each element. Also we are in the process of developing a method to identify answers to methodological questions within the elements and objective justification of a proposal draft.

We foresee an experiment that includes a pilot test with a control and experimental group of students. This test will provide an insight on whether there are improvements in students in the process of proposal drafting when helped by the intelligent tutor.

7. ACKNOWLEDGMENTS

We thank the three objective evaluators: Rene Castro Morales, Claudia I. Esquivel López, and J. Miguel García Gorrostieta. This research was supported by CONACYT, México, through the scholarship 1124002 for the first author. The second author was partially supported by SNI, México.

8. REFERENCES


