



Análisis Multinivel del Lenguaje en Textos de Estudiantes

Responsable

Dr. Samuel González-López¹

Dirección de Tecnologías de Información y Comunicación
Universidad Tecnológica de Nogales, Sonora
sgonzalez@utnogales.edu.mx

1. Antecedentes

La generación de conocimiento es una característica importante de las grandes potencias, las sociedades del conocimiento se dice son parte fundamental en el desarrollo de los países. México es un país que se encuentra en el desarrollo y mejora de su sistema educativo, así lo confirma la Reforma Educativa aprobada en diciembre de 2012 por el Gobierno Federal. La reforma establece que el Estado proveerá los materiales y métodos educativos, la organización escolar, la infraestructura educativa para el mejoramiento continuo y el máximo logro educativo de los estudiantes.

Uno de los principios relevantes en esta reforma es que todos los sectores tanto público como privado deberán colaborar para que la educación mejore y sea de mayor calidad². Bajo la dinámica de este principio de mejora y acorde a la reforma educativa se ha identificado un área de oportunidad en el nivel superior, la cual contempla el soporte en la redacción de documentos de estudiantes que están finalizando su programa educativo, específicamente en los documentos de tesis y propuestas de proyectos de investigación.

En México la investigación y desarrollo de nuevo conocimiento es soportado en su mayoría por los Centros de Investigación y algunas Universidades, con esquemas ya sea público o privado. El crecimiento del ingreso de alumnos a niveles posgrado ha sido ascendente, del 2004 al 2009 el incremento fue del 26%, aun así, existe un problema fuerte que afecta a todas las instituciones, el porcentaje de alumnos titulados. Este indicador a nivel posgrado se ubica alrededor del 48% según reportes de ANUIES. Cabe mencionar que en la mayoría de las instituciones que ofrecen posgrados la opción de titulación es a través de la realización de una tesis. A nivel Licenciatura este indicador para el periodo 2008-2009 fue de 79%, según datos de ANUIES. En este nivel los alumnos tienen

¹ *<http://utnogales.edu.mx/tic/cadesarrollosw/investigador.php>

² <http://www.presidencia.gob.mx/reformaeducativa/#sobre-la-reforma>

diferentes modalidades de egreso, una de ellas es la elaboración de una tesis. Para el nivel Técnico Superior Universitario el porcentaje de titulados en el mismo periodo que Licenciatura fue del 39%, en este nivel los alumnos culminan su programa con la elaboración de un documento tipo tesina.

Los factores o circunstancias para lograr la eficiencia terminal son diversos. En el estudio de Martínez, et al. [1] se presentan algunos factores de tipo administrativo y otros de tipo académico. En los factores de índole académico, los alumnos identificaron falta de asesoría, dificultad en la definición del tema a desarrollar, la elaboración del proyecto de tesis, entre otros. También concluyeron que la elaboración de una tesis es complicada para los estudiantes, debido a que no conocen con certeza las características de los elementos de la misma.

El proceso de elaboración de una tesis comienza con la estructuración de un borrador de propuesta o proyecto de investigación por parte del alumno. Comúnmente se involucran el asesor académico y el estudiante. Durante este proceso el asesor dedica tiempo a revisar el borrador que el alumno va construyendo y le proporciona sugerencias, esto se vuelve un ciclo hasta que se finaliza con un documento que cumple con características que se han establecido en libros sobre metodología de la investigación y guías institucionales. Este ciclo con frecuencia se vuelve lento, ya que algunas de las retroalimentaciones que genera el asesor se refirieren a la estructura de los elementos de la propuesta, por ejemplo la redacción que debe de llevar una hipótesis o un objetivo. Además, cada elemento de una propuesta de investigación posee características propias y que al integrarlos el resultado debe ser un documento coherente, es decir, que sus elementos estén conectados adecuadamente [2].

El léxico³ utilizado por los alumnos es una característica de todos los elementos de un borrador de propuesta, el cual debe ser considerado como una condición a cumplir en el documento final que el alumno entrega. En trabajo previo [3] se diseñó un analizador léxico tomando como base tres medidas léxicas: Variedad, Densidad y Sofisticación léxica. Como primer paso se evaluaron tesis de estudiantes de nivel Posgrado (Maestría y Doctorado) y se compararon con los de nivel Licenciatura y Técnico Superior Universitario (TSU). Las tesis fueron recolectadas de diferentes instituciones que publican las tesis en línea. Los resultados evidenciaron que los alumnos de posgrado tenían un léxico más variado y sofisticado que los estudiantes de menor nivel.

³ Vocabulario: conjunto de las palabras de un idioma, o de las que pertenecen al uso de una región, a una actividad determinada, a un campo semántico dado. <http://lema.rae.es/>

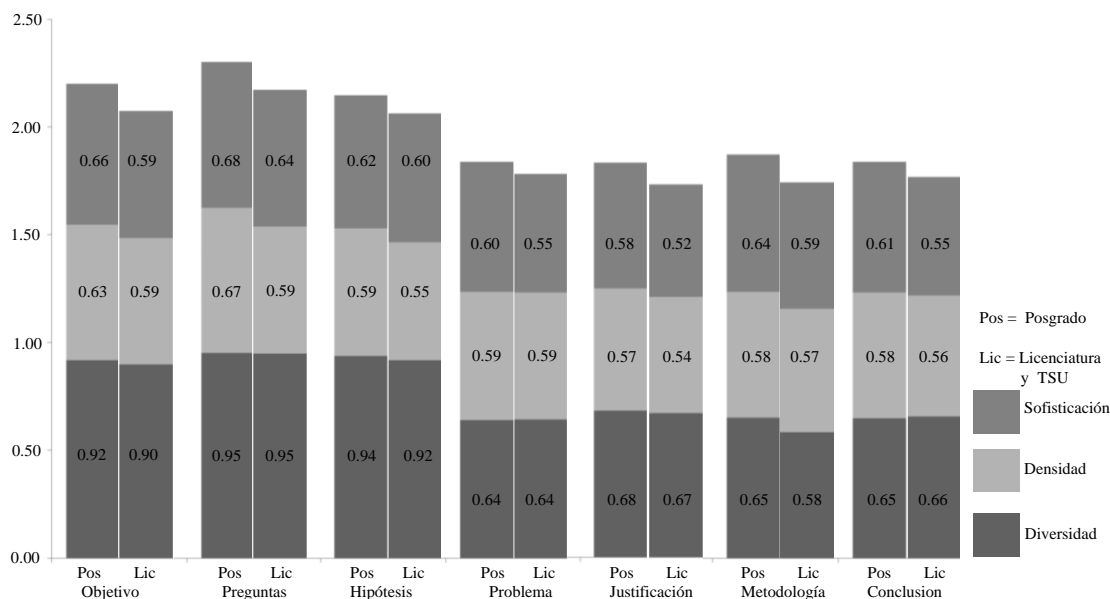


Figura 1. Resultados del Análisis Léxico

Se puede observar en la figura 1 que las barras del nivel Posgrado, obtuvieron un puntaje mayor comparado con los niveles Licenciatura y TSU. Por ejemplo para la sección *Objetivos* la sofisticación obtuvo un valor de 0.66 para Posgrado y 0.59 para Licenciatura y TSU. Lo cual indica que los estudiantes de posgrado tienen un uso de términos más enriquecido. La escala para las tres dimensiones va de 0 a 1. El tamaño de la colección analizada fue de 410 ejemplos. Una de las actividades pendiente a desarrollar es la implementación del analizador léxico bajo una plataforma web y en línea para que estudiantes puedan utilizarla.

Otra característica de las propuestas de investigación es estudiada en [4] y [5], donde se describen métodos para la evaluación de coherencia, un aspecto que toda propuesta de tesis tiene que satisfacer. Muchas definiciones de texto incluyen la coherencia como una característica necesaria. Una definición formal dada por Vilarnovo [6], establece que la coherencia de un texto es la conexión de todas sus partes en un todo; la interrelación de varios elementos de un texto. La coherencia en un borrador de propuesta de los estudiantes es importante porque si no está presente en cada uno de sus elementos, la idea central pierde sentido.

Los enfoques que se han abordado son el sintáctico y el semántico. El primero caracteriza el uso de una entidad en diferentes posiciones sintácticas y cómo se distribuyen entre oraciones adyacentes, mientras el semántico busca el grado de conexión temática entre las sentencias. Sin embargo la coherencia es solo un elemento de varios que un asesor revisa.

En el trabajo de Hernández y Ferreira [7], se evalúa la coherencia de una noticia del dominio policial en forma automática, es decir, dada una noticia policial escrita por un periodista, el sistema evaluador proporcionaba el grado de coherencia que la noticia tenía. En este trabajo se utilizó la técnica de Análisis Semántico Latente (LSA), primero se

recopiló un corpus del dominio de noticias policíacas el cual sirvió para entrenar al software y a partir de ese conjunto de textos, el evaluador medía la coherencia de la noticia. El resultado esperado era que el software evaluador se acercará a los resultados de evaluación que realizara un periodista y un profesor de español.

Un primer acercamiento a esta problemática en experimentos previos, fue el diseño de un analizador que evalúa la coherencia global [8]. En estos experimentos se aplicó la técnica de Análisis Semántico Latente a nuestro corpus recolectado, 380 elementos de entrenamiento y 80 de prueba, los cuales fueron etiquetados por anotadores con conocimiento del dominio de computación, considerando una escala de tres niveles para definir el nivel de coherencia del texto evaluado. La técnica LSA es un método matemático para el modelado y simulación por computadora del significado de las palabras y pasajes mediante el análisis de un corpus representativo del dominio [9].

El experimento consistió en construir un analizador de coherencia global con ejemplos considerados por los anotadores como de Alta coherencia, previamente procesados con la técnica de Análisis Semántico Latente. Posteriormente se evaluaron los 80 ejemplos de prueba y se contrastaron con el etiquetado de los Anotadores. Se realizó una prueba de acuerdo Fleiss Kappa entre anotadores y otra prueba entre anotadores y nuestro analizador - en este caso con Cohen Kappa (ver Tabla 1).

Kappa	Fleiss	Cohen
	Anotadores	Analizador de Coherencia
Alta	0,6862	0,0000
Media	-0,0378	0,2609
Baja	0,7353	0,4218
global	0,5458	0,2237

Tabla 1. Resultados de Prueba Kappa

Se alcanzó un nivel Considerable según Kappa, para los niveles de Alta y Baja coherencia entre anotadores, dándonos una base confiable de comparación. Los resultados entre el analizador no fueron cercanos a los anotadores en general, sin embargo para el nivel Bajo y Medio (los principales niveles a diagnosticar) fueron aceptables según Kappa⁴. Para el enfoque sintáctico de coherencia, planteamos en los objetivos específicos desarrollar un método que permita identificar las conexiones de los conceptos que se plasman en cada una de las secciones de una propuesta de investigación.

La estructura sintáctica de cada elemento de un proyecto de investigación es otra característica que se podría representar, es decir, la forma en que los alumnos construyen sus oraciones en cada uno de los elementos, por ejemplo los objetivos en su mayoría comienzan con un verbo en infinitivo, las preguntas de investigación siguen la estructura de una pregunta, esto según lo que establecen los autores de libros de metodología de la investigación y guías institucionales. Estas características sintácticas de cada elemento cobran importancia al momento que los alumnos redactan su proyecto de investigación.

⁴ Landis y Koch, 1977

Algunos trabajos han utilizado los modelos de lenguaje para caracterizar al texto, mayoritariamente en reconocimiento de voz donde ayudan por medio de probabilidades estimadas, a corregir ciertos errores que el discurso oral podría generar al transcribirse a texto [10].

Bajo este enfoque buscamos modelar las diferentes secciones de una propuesta extrayendo características de nuestros corpus con técnicas de procesamiento de lenguaje natural. Por ejemplo, se considera explorar en la sección de conclusiones aquellas oraciones que en ocasiones no corresponden (oraciones débiles), según los autores de metodología y guías institucionales, a una conclusión:

Oración débil:

En sí, la seguridad no debería de ser un problema, ni en las redes, ni en la vida cotidiana, pero como no se tiene una conciencia social, algunos humanos ya sea por avaricia pierden aquellos como generosidad.

Oración fuerte:

El nuevo sistema ayudará a reducir los costos de horas / hombre invertidas en el mantenimiento de la infraestructura.

Podemos observar que la oración débil se ajustaría mejor en una sección como instrucción o motivación. Mientras que la oración fuerte expresa un contraste positivo de la implementación de un sistema.

El uso de las técnicas de Procesamiento de Lenguaje Natural en nuestros experimentos previos nos brinda soporte para continuar trabajando con aquellas secciones que no hemos explorado. Este trabajo busca generar métodos que ayuden a la evaluación lingüística de ciertas características de los elementos de una propuesta de investigación, como el léxico, la coherencia, la estructura sintáctica propia que poseen los elementos, la identificación de elementos como importancia y necesidad, característicos de una Justificación. A continuación se muestra nuestra propuesta de solución:

1.1. Propuesta de solución:

En la parte baja de la figura 2 se encuentran los ocho elementos de una propuesta de investigación que se han considerado como elementos claves a evaluar: título, planteamiento del problema, justificación, objetivo, preguntas de investigación, hipótesis, metodología, y conclusiones. Se dará un tratamiento de los elementos a diferentes niveles, algunos serán procesados en los tres niveles y otros solo en algunos, esto debido a la naturaleza propia de los elementos. Los elementos que alcanzaran el cuarto nivel son objetivos y justificación, ya que son los que permiten en su construcción, responder a ciertas preguntas metodológicas guía. También se explorará la argumentación que debe darse principalmente en las secciones extensas (por ejemplo, planteamiento del problema, justificación y metodología) y buscar ser eficaces en la retroalimentación que se le provee al estudiante.

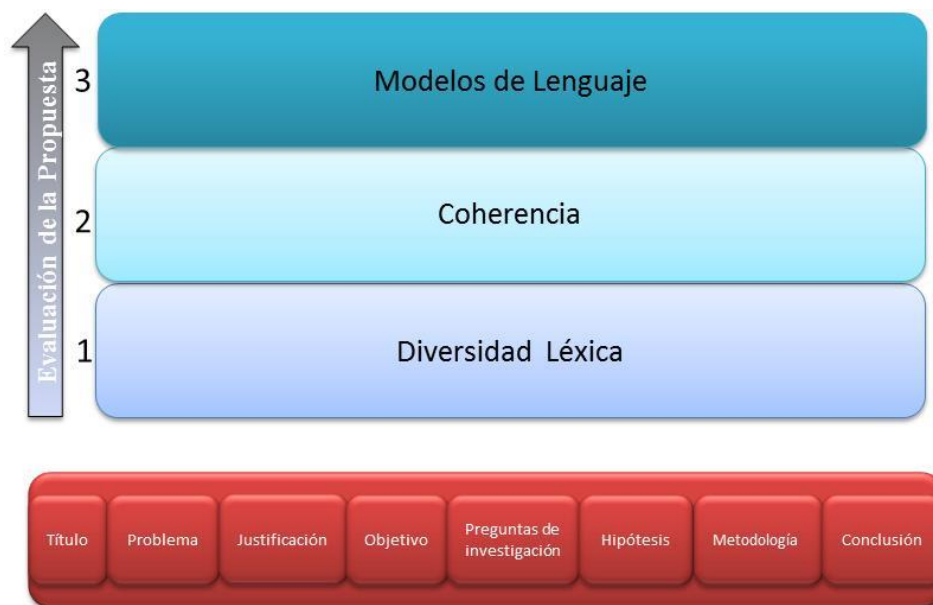


Figura 2. Diagrama de evaluación a tres niveles

El primer nivel busca determinar si el alumno está utilizando un vocabulario apropiado al área de computación y tecnologías de la información, además se desea determinar la riqueza del vocabulario, la cual indicará si el alumno tienen un manejo adecuado de la terminología del área, por ejemplo si un alumno escribe reiteradamente "sistema" en los elementos del proyecto, será indicio para sugerir que revise su léxico, buscando darle variedad. Este es el nivel por el cual comienza la evaluación y da una perspectiva a un nivel lingüístico básico. Actualmente este nivel ha sido finalizado y solo se plantea la implementación de un software en plataforma web para el acceso de cualquier estudiante que desee analizar su texto. Las publicaciones que soportan estos avances se detallan en resultados alcanzados.

El segundo nivel se enfoca en la evaluación de la coherencia a nivel local y global y se basa en la combinación de los dos aspectos el semántico y el sintáctico, dado que busca capturar si los elementos de la propuesta de investigación son coherentes semánticamente al área de computación y tecnologías de la información, pero también busca que los elementos por sí mismo sean coherentes, aquí es donde ese incorpora el enfoque sintáctico. La combinación de los dos aspectos se basa en estudios previos, donde se ha demostrado que las técnicas que se utilizan capturan aspectos complementarios de la coherencia [11]. En este nivel no se evalúa el elemento título, por la baja relevancia que pudiera tener. Actualmente hemos concluido el aspecto de Coherencia global y estamos trabajando en el aspecto de Coherencia local. Las publicaciones donde se documentan estos avances se incluyen en resultados alcanzados.

En el tercer nivel se busca atacar a la tarea de evaluación de un borrador de propuesta, capturando en modelos de lenguaje la estructura sintáctica que guarda cada elemento, este tercer nivel busca una independencia temática, es decir, se desea capturar qué tipo de

elementos del discurso están siendo utilizados y cómo son utilizados. Por ejemplo el uso de verbos, adverbios, sustantivos. Este nivel se encuentra por arriba de lo temático ya que queda en segundo plano el contenido y se enfatiza en los elementos sintácticos, al hacer esto se busca caracterizar cada elemento y definir una estructura o patrón sintáctico. Este nivel se ha empezado a explorar con experimentos a nivel de modelos de lenguaje.

Un aspecto importante del trabajo que se ha venido desarrollando y de la propuesta es que se basa en una colección de documentos de tesis y propuestas a distintos niveles académicos (técnico superior universitario, licenciatura, maestría y doctorado). Estos documentos han sido revisados en algún momento por un grupo de profesores y nos sirven tanto para identificar (minar) las características de interés, así como para evaluar los métodos.

La evaluación a tres niveles daría una perspectiva global de la estructura del borrador de propuesta al alumno y permitiría que antes de llegar con el asesor académico, la propuesta muestre determinadas características que las guías institucionales y los autores de metodología establecen. Por otro lado permitiría al asesor dedicar mayor tiempo a la revisión del contenido. Este trabajo busca integrar las evaluaciones de cada uno de los niveles, de tal forma que formen una plataforma de evaluación lingüística-estructural.

1.2. Resultados previos alcanzados:

- ✓ Samuel González López and Aurelio López-López. 2012. Supporting the review of student proposal drafts in information technologies. In Proceedings of the 13th annual conference on Information technology education (SIGITE '12). ACM, New York, NY, USA, pp. 215-220.
- ✓ Jesús Miguel García Gorrostieta, Samuel González López, Aurelio López-López. 2012. Assessing and Advising on Lexical Richness in an Intelligent Tutoring System. Research in Computing Science, pages 29-36, Vol. 56.
- ✓ Samuel González López and Aurelio López-López. 2014. Mining Domain Knowledge for Coherence Assessment of Students Proposal Drafts. Chapter in Educational Data Mining, Studies in Computational Intelligence. Springer International Publishing, pp. 229-255.
- ✓ González-López, Samuel; López-López, Aurelio Mining of Conclusions of Student Texts for Automatic Assessment , The 28th International FLAIRS2015 Conference May 18 - 20, Hollywood, Florida, USA, 2015.

2. Hipótesis

El análisis y evaluación a distintos niveles permitirá la evaluación de las principales características en los elementos de un borrador de propuesta de investigación, que posibilitará el proveer a los estudiantes una retroalimentación en etapas tempranas del desarrollo de la propuesta o tesis.

3. Objetivo general

Diseñar y desarrollar métodos computacionales para analizar lingüísticamente a distintos niveles borradores de propuestas de estudiantes, de Técnico Superior Universitario y

Licenciatura, alcanzando desempeños comparables a revisores humanos, con el fin de sentar las bases para el apoyo a la escritura en etapas tempranas de elaboración.

Objetivo específicos:

- Explotar recursos existentes (tesis, tesinas y propuestas de investigación) a nivel posgrado y licenciatura, para analizar nuevos textos.
- Implementar un analizador léxico en línea, para el acceso de estudiantes con la finalidad de ayudarles en la redacción de su propuesta.
- Implementar una prueba piloto con un grupo de control y uno de experimentación, para el validar el analizador léxico.
- Construir modelos de lenguaje que caractericen cada elemento de una propuesta, permitiendo generar patrones sintácticos para analizar nuevos textos.
- Definir métodos para identificar posibles relaciones entre los elementos de una tesis, por ejemplo, la conexión de los objetivos con las conclusiones.

4. Metas: Científicas y de Formación de Recursos Humanos

En esta investigación se busca conformar un grupo de trabajo bajo la línea de investigación, la cual gira en torno al análisis de propuestas de investigación de estudiantes de licenciatura y Técnico Superior Universitario, para contribuir con nuevos conocimientos.

También se contempla fortalecer la colaboración con centros de investigación como el Instituto Nacional de Astrofísica, Óptica y Electrónica.

4.1. Metas Científicas

- Contribuir con métodos innovadores en Aprendizaje Mejorado por Tecnología (Technology Enhanced Learning - TEL), explotando el texto como materia prima para los desarrollos y experimentos.
- Aportar nuevas técnicas en Procesamiento de Lenguaje Natural y Lingüística Computacional para el análisis de lenguaje escrito en Español.
- Impulsar el estudio de documentos de propuestas de investigación y tesis, para entender cómo se manifiestan diversos fenómenos del lenguaje tales como riqueza léxica, coherencia, legibilidad, planteamiento metodológico, argumentación, entre otros.
- Incursionar en la investigación de nuevos mecanismos que guíen, retroalimenten y mantengan el interés de los estudiantes en mejorar su escritura.
- Formar un grupo especializado en estas técnicas y líneas de investigación.
- Contribuir a la comunidad de investigación con recursos para posteriores investigaciones, tanto en recopilación, como en documentos etiquetados por instructores con características específicas.
- Llevar a cabo evaluaciones de los métodos formulados tanto en una colección experimental recopilada como en grupos de estudiantes.

4.2. Metas de formación de recursos humanos y difusión de la investigación

- Formación de recursos humanos especializados en las temáticas del proyecto a nivel licenciatura. Se busca que los estudiantes de Licenciatura se interesen en el tema y continúen a nivel maestría con sus estudios.
- Desarrollo de una aplicación en línea que incorpore los métodos formulados con el fin de que apoye a estudiantes en la formulación de sus propuestas de investigación o tesis, y de manera colateral apoye a instructores en la revisión de borradores.
- Difusión de los resultados obtenidos del proyecto de investigación entre la comunidad de estudiantes y profesores universitarios a través de publicaciones.
- Impulsar a la comunidad interesada en las temáticas del proyecto por medio de un foro a realizarse en el marco de las Jornadas académicas de Tecnologías de la Información 2017. Este evento aglomera a todas las Universidades Tecnológicas del Estado de Sonora.

Asociado a estas metas, se contempla la titulación de 1 estudiante de nivel licenciatura.

4.3. Metas de derivación tecnológica

- Desarrollo de una aplicación web que incorpore los métodos formulados con el fin de apoyar a estudiantes en la formulación de sus propuestas de investigación o tesis, la cual se registrara con derechos de autor. Específicamente el análisis léxico.
- Desarrollo de herramientas para apoyar en la revisión (etiquetado) de los documentos de interés.

5. Metodología

Para alcanzar los objetivos planeados, proponemos llevar una evaluación a tres niveles: el primer nivel se refiere a la evaluación del léxico, el segundo se enfoca sobre la evaluación de la coherencia global y local, el tercer nivel se centra en modelos de lenguaje que caractericen a los elementos de un borrador. Cada nivel requiere técnicas específicas, dada la naturaleza variada de las características a analizar. A continuación se define la metodología a utilizar en este proyecto para alcanzar el cuarto nivel de nuestra solución propuesta.

1. Incrementar el tamaño del corpus considerando las propuestas de proyectos de investigación y tesis que permitan identificar características de interés. Se consideran tesis en español del área de computación y tecnologías de la información. El tamaño actual de nuestra colección asciende a 300 tesis.
2. Diseñar e implementar un analizador léxico en línea, utilizando los resultados previos de nuestros experimentos.
3. Realizar una prueba piloto con estudiantes de nivel licenciatura (grupo de control y experimentación) para validar si el analizador léxico ayuda al estudiante a mejorar su escritura.

4. Diseñar e implementar modelos de lenguaje que permitan caracterizar los componentes sintácticos de cada elemento de una propuesta de investigación. La construcción de estos modelos contará con las siguientes etapas:
 - a. Adecuar el corpus para el entrenamiento.
 - b. Aplicar técnicas para construir modelos de cada elemento de una propuesta de investigación, tales como clasificación, sumarización, agrupamiento y modelos de lenguaje.
 - c. Generar una escala del grado o nivel de cercanía entre los modelos generados y los modelos de cada elemento analizado.
 - d. Generar un conjunto de referencia usando anotadores, para posteriormente realizar la prueba de Kappa para medir el nivel de acuerdo entre los anotadores y el método de análisis.
5. Diseñar un método para identificar las frases (patrones) que manifiesten la importancia y necesidad de una Justificación.
6. Diseñar una evaluación que permita comparar los resultados obtenidos por el analizador de coherencia y los modelos de lenguaje con juicios humanos.
 - a. Para ello se plantea apoyarnos en profesores que puedan evaluar un proyecto de investigación desde los enfoques coherencia, estructural-sintáctica y de respuesta a preguntas complejas y comparar los resultados con cada una de los analizadores.
7. Desarrollar mecanismos de retroalimentación al estudiante, así como de motivación para seguir mejorando su texto. Se planea explorar estrategia de juegos serios, es decir juegos que tienen otros fines más allá de la diversión, en este caso la educación.

6. Grupo de trabajo

6.1. Instituciones participantes

Los investigadores forman parte de la Universidad Tecnológica de Nogales Sonora y del Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOE) ubicado en Tonantzintla Puebla.

Los estudiantes de Licenciatura forman parte de la carrera de Tecnologías de la Información de la Universidad Tecnológica de Nogales Sonora.

6.2. Integrantes

- Doctor Samuel González López: obtuvo su doctorado en Ciencias en el área de Ciencias Computacionales en el Instituto Nacional de Astrofísica, Óptica y Electrónica, Tonantzintla Puebla, México. También estudió la maestría en Ciencias en Ciencias de la Computación y la maestría en Ciencias de Ingeniería Industrial en el Instituto Tecnológico de Hermosillo, Sonora, México. Actualmente es profesor investigador en la Universidad Tecnológica de Nogales, Sonora.

- Dr. Aurelio López López (Investigador Principal): es Profesor-Investigador Titular B del Instituto Nacional de Astrofísica, Óptica y Electrónica. Obtuvo su doctorado en Ciencias Computacionales y de la Información en Syracuse University, Syracuse Nueva York, E.U.A. Sus áreas de interés son representación del conocimiento, extracción y recuperación de información de textos, minería de textos, así como tratamiento de lenguaje natural. Forma parte del Laboratorio de Tecnologías de Lenguaje del Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOE). Además pertenece al Sistema Nacional de Investigadores (SNI) nivel I.
- M.C. Indelfonso Rodríguez Espinoza: egresado de la Universidad de Occidente en la ciudad natal Guasave, Sinaloa, así mismo título de la Maestría en Sistemas Computacionales en el Instituto Tecnológico de Nogales, Sonora y en proceso de titulación de la Maestría en Educación: Campo Formación Docente de la Universidad Pedagógica Nacional campus Nogales, Sonora. Experiencia profesional en la industria maquiladora, específicamente en el área de Tecnologías de Información (TI) durante 8 años, además 16 años como catedrático en la Universidad Tecnológica de Nogales, Sonora.
- TSU Fernando Wirichaga Primero: Estudiante de Ingeniería que participaría en este proyecto. El estudiante tiene experiencia con algunas técnicas de inteligencia artificial.

7. Infraestructura disponible

- La dirección de Tecnologías de la Información y Comunicación de la Universidad cuenta con cubículos personales. Para el desarrollo del proyecto se cuenta con un servidor T410 Power Edge en el cual se alojan documentos del Cuerpo. También se cuenta con acceso a Internet, pero es necesario el acceso a bases de datos externas para descargar artículos arbitrados como: ACM, Springer Link, IEEE.
- Actualmente la Universidad Tecnológica de Nogales cuenta con laboratorios para la impartición de las clases de programa. Sin embargo, se plantea la adquisición de equipo para establecer un Laboratorio de Inteligencia Artificial que permita a los integrantes del equipo de trabajo desarrollar cada una de las actividades. Este laboratorio serviría para futuros proyectos del Cuerpo Académico. Actualmente la Universidad cuenta con espacio disponible para su instalación.

8. Resultados entregables esperados

Artículos Científicos: Se plantea la publicación de al menos 3 artículos: 1 artículo en revista indizada, un artículo en Congreso Internacional y un artículo en congreso Nacional. Una de revista considerada es Technology, Knowledge and Learning (Indizada en SCOPUS).

Los eventos que se han considerado para someter y publicar los artículos son:

- ✓ Mis4Tel: Methodologies and Intelligent Systems for Technology Enhanced Learning. Editorial Springer.
- ✓ EC-TEL: The European Conference on Technology Enhanced Learning. Editorial Springer.
- ✓ FLAIRS: The Florida Artificial Intelligence Research Society. Editorial AAAI.
- ✓ WILE: Workshop on Intelligent Learning Environments
- ✓ CIRC Congreso Internacional de Computación

Estas conferencias son de carácter internacional y de amplia distribución a través de sus editoriales. Además, estos eventos tienen tópicos muy relacionados a nuestro trabajo.

Alumnos graduados: Formar estudiantes a través del proyecto es una meta a seguir, por esta razón se tiene programado graduar a:

- ✓ Un estudiante de nivel Licenciatura: Wirichaga Primero Fernando
- **Base de Datos:** Se contempla la puesta en línea del corpus recolectado a través de una Base de Datos y un sitio Web. Además se espera poner a disposición de otros investigadores los corpus que sean etiquetados por anotadores (asesores académicos).
- **Foro de Difusión:** Con la finalidad de difundir los resultados y promover nuestras herramientas entre los estudiantes, se desarrollará un foro de difusión donde se presenten nuestros avances y productos (de las diferentes secciones de un borrador de propuesta). Así como, las líneas de investigación que se han derivado del proyecto. También se planea invitar a investigadores que estén realizando trabajos con documentos de estudiantes. Con el desarrollo del foro se espera que investigadores del área se interesen en las líneas presentadas y se incrementen los trabajos relacionados a mejorar la escritura de los estudiantes.
- **Reportes Técnicos.** Se contempla también la generación de reportes técnicos a lo largo de la duración del proyecto donde se documenten más ampliamente los métodos, experimentos y resultados.

9. Bibliografía

- [1] Martínez, J., Gutiérrez, D., Hernández F. 2007. Problematic terminal efficiency developed graduate programs in distance mode, in the IPN. International Congress of Educational Innovation.
- [2] Sampieri, R. 2006. Metodología de la Investigación. México DF, Mc Graw Hill.
- [3] Samuel González López and Aurelio López-López. 2012. Supporting the review of student proposal drafts in information technologies. In Proceedings of the 13th annual conference on Information technology education (SIGITE '12). ACM, New York, NY, USA, pages 215-220.

- [4] Foltz, P., Kintsch, W., Launder, T., 1998. Textual Coherence using Latent Semantic Analysis. Colorado USA: Discourse Processes, 285-307
- [5] Elsner M., and Charniak E. 2008. Coreference-inspired Coherence Modeling. Proceedings of the Association for Computational Linguistics: Human Language Technologies (ACL-HLT 2008), 41-44.
- [6] Vilarnovo, A. 1990. Text Coherence: Internal Coherence or External Coherence? ELUA Journal, 229-239.
- [7] Hernández, S. and Ferreira, A. A. 2010. Evaluación automática de Coherencia textual en noticias policiales utilizando Análisis Semántico Latente. Revista de Lingüística Teórica y Aplicada, 48(2): pp. 115-139.
- [8] Samuel González López and Aurelio López-López. 2014. Mining Domain Knowledge for Coherence Assessment of Students Proposal Drafts. Book Chapter in Educational Data Mining, Studies in Computational Intelligence. Springer International Publishing, pages 229-255.
- [9] Dumais, Thomas K Landauer and Susan. Latent semantic analysis. s.l. : Scholarpedia, 2008. Revisión #90365.
- [10] Kuan-Yu Chen and Berlin Chen. 2011. Relevance Language Modeling For Speech Recognition. International Conference on Acoustics, Speech, and Signal Processing. ICASSP'11, 5568-5571.
- [11] Lapata, M., Barzilay, R. 2005. Automatic Evaluation of Text Coherence: Models and Representation. In Proceedings of International Joint Conference on Artificial Intelligence, 1085-1090.