



Revista de Investigación Educativa 11

julio-diciembre, 2010 | ISSN 1870-5308 | Xalapa, Veracruz

© Todos los Derechos Reservados

Instituto de Investigaciones en Educación | Universidad Veracruzana

Contribución del Análisis Estadístico Textual a la labor del tutor académico y el orientador educativo

Ma. de los Ángeles Silva Mar

Académica de la Facultad de Pedagogía
Universidad Veracruzana, Región Poza Rica-Tuxpan

María Luisa Hernández Maldonado

Académica de la Facultad de Estadística e Informática
Universidad Veracruzana

Areli Sosa Mora

Colaboradora de la Facultad de Estadística e Informática
Universidad Veracruzana

Este trabajo es parte del resultado del análisis textual realizado a las respuestas libres contenidas en la encuesta sobre “Búsqueda de la personalidad de los estudiantes” que la Universidad Veracruzana aplicó a parte de su comunidad estudiantil. Entre los objetivos que se plantean se encuentra el estudio sobre el autoconocimiento que los estudiantes manifiestan tener. Este trabajo contribuye a la labor del Tutor Académico y al quehacer del Orientador Educativo que, entre otras, tienen como finalidad el reforzar y apoyar a los estudiantes en sus requerimientos para mejorar su rendimiento escolar.

Para el tratamiento de este instrumento, cuyas variables son de naturaleza textual, se ha realizado un análisis estadístico de textos. Este análisis ha integrado distintas técnicas y métodos (combina el análisis multidimensional de datos y la lexicometría) que facilitan la tarea en el análisis de encuestas que contienen preguntas de respuesta libre. En una primera etapa, se aplican procedimientos que permiten obtener el número de ocurrencias de las unidades verbales básicas (generalmente palabras); posteriormente, a partir de esos recuentos es posible realizar un análisis factorial de correspondencias que muestre los rasgos estructurales que los individuos manifiestan a través de sus respuestas.

Palabras clave: Análisis estadístico de textos, Análisis de correspondencias, Orientación educativa y Tutor Académico.

Recibido: 11 de enero de 2010 | **Aceptado:** 30 de marzo de 2010

Para citar este artículo:

Silva, M. A., Hernández, M. L. & Sosa, A. (2010, julio-diciembre). Contribución del Análisis Estadístico Textual a la labor del tutor académico y el orientador educativo. *CPU-e, Revista de Investigación Educativa*, 11. Recuperado el [fecha de consulta], de <http://www.uv.mx/cpue/num11/opinion/silva-analisis-estadistico.html>

Contribución del Análisis Estadístico Textual a la labor del Tutor Académico y el Orientador Educativo

Introducción

Los hábitos y actitudes hacia el estudio han sido motivo de investigación desde hace muchos años; se sabe que existen correlaciones significativas entre hábitos y actitudes hacia el estudio y las calificaciones de los estudiantes. Estudiantes con puntuación alta en una prueba de hábitos de estudio obtienen éxito en el trabajo escolar y viceversa, estudiantes que no presentan buenos hábitos y métodos de estudio, tienen un rendimiento deficiente. Desde hace varias décadas se sabe que las principales dificultades que los estudiantes enfrentan se refieren a técnicas y hábitos de estudio principalmente en las áreas de: memorización como técnica para aprender, distribución del tiempo y problemas de concentración.

Así, el aprender a estudiar y mejorar sus hábitos y técnicas para lograrlo, constituye un arduo trabajo para los estudiantes y un importante reto para las instituciones educativas. El estudiante que tiene dificultad para afrontar con éxito las exigencias que el sistema le plantea, requiere mejorar la habilidad de aprender, a fin de lograr un mejor nivel educativo y una formación académica que permita un desempeño profesional más eficiente.

La Universidad Veracruzana, atendiendo esta necesidad, en febrero de 2009 aplicó una encuesta a 358 estudiantes inscritos en la Facultad de Pedagogía, Región Poza Rica-Tuxpan. La encuesta estaba integrada por tres cuestionarios: uno que contempla datos generales del estudiante, otro de 70 preguntas orientadas a definir las siete áreas de hábitos de estudio, y uno más, que es un test de frases incompletas. Los datos fueron recolectados mediante el Sistema de Cuestionarios Móviles ubicado en una página web.¹ Cada estudiante entró a la página y fue contestando el cuestionario generando así la base de datos con la cual se trabaja. Respecto a los hábitos de estudio, conforme los datos entraron el programa automáticamente generó, para cada estudiante, un informe sobre sus hábitos, adecuados o inadecuados, tomando en cuenta las distintas áreas bajo

1. <http://sapp.uv.mx/cuestmovil/cuestionario.aspx?idcuestionario=Habito32009094831>

estudio. El informe respectivo fue enviado por correo electrónico a cada estudiante y al correspondiente tutor académico.

Sobre el test de frases incompletas² es una prueba proyectiva que mide los aspectos del individuo, en cuanto a su familia, género, relaciones interpersonales así como el autoconcepto. El test consta de once frases incompletas que el estudiante, de acuerdo con las instrucciones, contestó con lo primero que se le vino a la mente, lo cual se espera que refleje su propio autoconocimiento y grado de autorrealización. Las respuestas a las frases incompletas fueron tratadas como variables de tipo textual conjuntamente con otro tipo de variables que fueron tratadas como ilustrativas.

El test consta de 11 frases incompletas, en este trabajo se presentan resultados del análisis hecho a cuatro de esas frases.

A través de estas frases incompletas se pretende que el orientador-tutor conozca las capacidades, habilidades, los valores y las metas que posee cada estudiante de Pedagogía, ya que a través de sus respuestas puede identificar aspectos relevantes para realizar su función de orientador en cuanto al desempeño en el proceso enseñanza-aprendizaje, decisión de su profesión en función de su vocación, proyecto de vida y carrera, entre otros asuntos que todo orientador-tutor puede abordar con sus tutorados.

Las respuestas libres son tratadas a través de un Análisis Estadístico para Datos Textuales (AEDT) que permite explorar desde el significado de un texto, lo que las personas quieren realmente decir cuando responden libremente. Los datos que se manejan en este tipo de respuestas son de naturaleza textual y se utiliza toda la información disponible sobre el encuestado, incluyendo las respuestas a preguntas cerradas. El análisis realizado se hace sin operaciones de codificación previas. El AEDT puede estar al principio de un proceso de investigación, puede ser el punto de partida para tratar un problema (Bécue, 1991). Sus resultados pueden despertar nuevas interrogantes o servir para profundizar en la investigación.

El análisis estadístico para datos de naturaleza textual

2. Frases incompletas incluidas en el test: 1. Razones por las cuales elegí la carrera... 2. Soy muy bueno a la hora de..., 3. Me quiero a mí mismo(a) porque..., 4. Consigo éxito en lo que me propongo porque..., 5. No me asusta..., 6. Estoy perfectamente dotado(a) para..., 7. Sé que conseguiré lo que quiero porque..., 8. No siento miedo ante..., 9. Me gusta mi carrera pero..., 10. Mi rendimiento escolar lo considero..., 11. Lo que más deseo en la vida...

Los métodos y técnicas estadísticas implican procedimientos para contar las ocurrencias de las unidades verbales básicas que generalmente son palabras; la información que se deriva a partir de esos resultados permite realizar un análisis desde una perspectiva multidimensional que sirve para poner de manifiesto rasgos estructurales de los individuos (Abascal & Franco, 2002).

Lo primero que es necesario realizar es la preparación del corpus.³ De acuerdo con los objetivos de estudio, conviene que el investigador depure y prepare la base de datos definiendo los procedimientos apropiados para su registro y la depuración del fichero antes de iniciar el análisis. Es recomendable realizar la “normalización del corpus”, donde se detectan las palabras como formas gráficas, se elimina la ambigüedad, si es conveniente se realiza una serie de cadenas reconocidas como nombres propios (de personas, de lugares, etc.), se convierten las secuencias de formas gráficas reconocidas como multipalabras en cadenas unitarias, para utilizarlas como tales durante el proceso de análisis (por ejemplo: “en otras palabras” se transforma en “en_otras_palabras”).

Posteriormente, de acuerdo con las necesidades de estudio y el criterio del investigador, se define la segmentación de contextos elementales; en el caso del análisis a un conjunto de respuestas libres, cada respuesta de un individuo es considerada como un segmento del corpus. En otro tipo de corpus, como en el análisis de documentos (entrevistas, páginas web, sentencias jurídicas, discursos, libros, etc.) se realizan cortes en segmentos (frases, párrafos y/o textos breves) que faciliten la exploración de los datos.

El análisis textual inicia con el estudio de las unidades estadísticas: se identifican las formas gráficas que son en general letras (unidad con una secuencia de caracteres no delimitadores comprendida entre dos caracteres delimitadores, que pueden ser espacios en blanco o signos de puntuación) y segmentos repetidos que son una secuencia de dos o más formas gráficas o palabras que aparecen más de una vez en el corpus que se analiza. El conjunto de formas gráficas de un texto es lo que constituye su vocabulario (Lebart & Salem, 1989).

El análisis inicia con procedimientos que implican contar las ocurrencias de las unidades verbales básicas (palabras), se analizan y depuran los listados que se producen. Para el estudio de la riqueza de vocabulario, se procede a hacer un filtro de las palabras expresadas por los individuos, se eligen las que presentan frecuencias altas (queda a criterio del investigador determinar la frecuencia). Así

3. El corpus es una colección de uno o más textos seleccionados para el análisis. En este estudio es el conjunto de respuestas a cada pregunta abierta.

mismo, es importante identificar la presencia de palabras herramienta (preposiciones, artículos, conjunciones, pronombres...) y qué tan oportuno es tomarlas o no en cuenta, para la interpretación. Si su reparto es aleatorio y se aproximan o se sitúan en la parte central del gráfico, interesa eliminarlas. Por último, la presencia de varias formas de un mismo verbo puede justificar la lematización del corpus, que consiste, justamente, en identificar como equivalentes las formas gráficas que tienen la misma raíz y un significado semejante. La lematización del corpus no es indispensable e incluso puede ocasionar problemas (Lebart, Morineau & Bécue, 1989).

Además de las palabras, si consideramos que un dato textual debe reflejar lo que las formas empleadas (palabras) quieren decir, el análisis toma como otra unidad de estudio los “segmentos de texto con sentido”; es decir, probablemente las palabras solas no reflejan el sentido correcto con el que fueron empleadas, una frase rectifica la idea o sentido del contexto al que se refieren. Por ejemplo, la frase “No me gusta mi carrera” contradice la idea de “Me gusta mi carrera pero”.

Después de tener el vocabulario completo y depurado, se procede a estudiar la información derivada de estos recuentos. Con este vocabulario se construye una tabla léxica y se aplica un Análisis factorial de Correspondencias (AC).

Tablas léxicas

El análisis de correspondencias aplicado a una matriz de “documentos x términos” es un método muy eficiente para visualizar la similitud semántica entre términos o palabras (Benzécri, 1976; Lebart, Salem & Berry, 1998; Lebart, Morineau & Piron, 2000), semejanzas entre documentos o textos y asociaciones entre términos y documentos. Cada corpus descrito se representa por una matriz con I filas y J columnas, en donde la i -ésima fila contiene el vector cuyos J componentes son las frecuencias de las J palabras (o términos) en el i -ésimo documento. Esta tabla se llama tabla léxica o matriz “documentos x términos” (Hernández, 2007).

En el caso de respuestas a preguntas abiertas, para construir la tabla léxica se considera: identificar las palabras distintas o segmentos y contar su frecuencia absoluta en las distintas partes del corpus. A partir de las unidades lexicales y el valor numérico de su frecuencia en los segmentos, el investigador analiza la importancia de incluir o excluir determinadas palabras. Una forma de realizar

esta depuración es observando las palabras de más baja o alta frecuencia (estas son las que normalmente se eliminan) ya que generalmente las de menor frecuencia estadísticamente no son significativas y las de mayor frecuencia comúnmente son palabras ambiguas que no enriquecen el análisis, porque sólo funcionan como conectores, artículos, preposiciones, etc.

Una tabla léxica agregada es una tabla de contingencia cuyo término general recoge el número de veces que la forma i ha sido utilizada en la respuesta libre, por el conjunto de individuos que eligen la modalidad j en una respuesta cerrada. A partir de esta tabla que cruza palabras x grupos de respuestas, se pueden comparar los perfiles léxicos de los segmentos de población definidos (Abascal, 2002). Se puede obtener una tabla léxica agregada para cada pregunta cerrada, y comparar los perfiles léxicos de las diferentes categorías de la población.

Análisis de Correspondencias (AC) de la tabla léxica

El AC es un método descriptivo clásico para el estudio de tablas de contingencia. Describe la asociación entre las dos variables categóricas correspondientes a la tabla de contingencia. Se puede ver, también, como la búsqueda de la mejor representación simultánea de las modalidades de las dos variables categóricas y estudia las relaciones eventuales existentes entre ellas (Lebart, 1984).

Los principios teóricos del AC se remontan a los estudios sobre tablas de contingencia hechos por Fisher en 1940. Entre sus principales precursores figuran Guttman (1941) y Hayashi (1956), Escofier y Cordier (1965) y Benzécri (1973), quienes privilegian las propiedades algebraicas y geométricas del método.

En el AEDT el AC identifica las palabras que contribuyen a la inercia de los diferentes ejes seleccionados. Dos conjuntos de palabras están asociados a un mismo eje factorial, uno a la parte positiva, el otro a la parte negativa. Una misma palabra puede estar asociada a diferentes ejes factoriales, dependiendo del contexto puede pertenecer a diferentes conjuntos (Hernández, 2007).

El AC de una tabla léxica proporciona una visión gráfica, simplifica la información que contiene la tabla y pone en evidencia las diferencias entre los distintos perfiles léxicos: las representaciones permiten situar a los individuos en un espacio determinado por las formas; es posible comparar los perfiles de sus respuestas, localizando aquellos más semejantes y las formas causantes de estas semejanzas. Del mismo modo, se detectan los individuos más diferentes y se trata de explicar las formas que causan esas diferencias. En el caso de una

tabla léxica agregada, se pueden estudiar las posiciones relativas de diferentes grupos de individuos en relación al vocabulario empleado y caracterizarlos por las palabras utilizadas.

La representación de las formas permite estudiar las proximidades entre aquellas formas que son utilizadas simultáneamente por los mismos individuos, es decir, estudiar los contextos, ya que la proximidad entre dos formas es mayor cuando aparecen en la misma respuesta.

El análisis de datos textuales se puede complementar con otros métodos lexicométricos y técnicas clásicas como el estudio del vocabulario, las concordancias, las formas características, etc. (Lebart & Salem, 1988; Bécue, 1991)

Aplicación del AEDT en el estudio sobre la personalidad del estudiante

Se realiza un análisis textual a las respuestas de las frases incompletas contenidas en la encuesta. Se presentan los resultados del análisis hecho a cuatro frases que califican el área de autoconocimiento del estudiante. Se analiza cómo el estudiante refleja su propio autoconocimiento. Para el tratamiento de los cuatro corpus se utiliza el paquete de programas de SPAD-T (Lebart *et al.*, 1989).

Los resultados que se presentan corresponden al AEDT hecho a las siguientes primeras tres frases incompletas; para complementar la información se presentan los resultados de la cuarta frase:

1. Soy muy bueno (a) a la hora de ...
2. Me quiero a mí mismo(a) porque ...
3. No me asusta ...
4. Mi rendimiento escolar lo considero ...

En cada análisis se consideran únicamente las palabras de frecuencia mayor o igual a 3 y se eliminan las palabras herramientas identificadas (*lo, los, la, las, un, una, de, el, en, para, por, porque, etc.*).

Análisis y depuración del listado de palabras

La tabla 1 presenta las estadísticas del corpus. La frase “Me quiero a mí mismo(a) porque...” presenta en sus respuestas el vocabulario más amplio (1730 palabras, de las que el 17.7% son distintas), comparado con las respuestas de las otras dos frases. Después del filtro el vocabulario se reduce a 770 palabras (63 distintas), se conserva el 45% de palabras para el análisis.

Tabla 1. Estadísticas del corpus para la selección de datos

	1. Soy muy bueno(a) a la hora de...	2. Me quiero a mí mismo(a) porque...	3. No me asusta...
Número de Estudiantes	358	358	358
Total de palabras	851	1730	1084
No. de palabras distintas	248	304	320
% palabras distintas	29.1%	17.7%	29.5%
Estadísticas después del filtro. Frecuencias de palabras: >ó =3			
Total de palabras	416	770	351
No. de palabras distintas repetidas	53	63	51
% a tener en cuenta	49%	45%	32%

De las respuestas obtenidas de la frase “Soy muy bueno(a) a la hora de...” se desprende que los estudiantes utilizan un vocabulario más concreto (851 palabras utilizadas); de las palabras que utilizaron, después del filtro, queda (416 palabras el 49%) a tomar en cuenta para el análisis. Por último, en las respuestas de “No me asusta...” los estudiantes utilizaron 1084 palabras (el 29.5% son palabras distintas), después del filtro el corpus se redujo a 351 palabras; 51 de ellas son distintas y se trabaja con el 32% del vocabulario empleado (ver tabla 1).

La tabla 2 muestra una parte del vocabulario empleado por los estudiantes en cada una de las frases incompletas. Se realiza posteriormente un análisis de correspondencias para cada uno de los tres corpus, su representación e interpretación se presentan por separado.

Tabla 2. Muestra del vocabulario empleado por los estudiantes en cada una de las frases incompletas

1. Soy muy bueno(a) a la hora de...		2. Me quiero a mí mismo(a) porque...		3. No me asusta...	
Palabras empleadas	Frecuencia	Palabras empleadas	Frecuencia	Palabras empleadas	Frecuencia
HACER	42	SOY	232	NADA	43
ESTUDIAR	34	PERSONA	37	RETOS	39
EXPONER	21	GUSTA	34	MUERTE	16
LEER	21	ÚNICA	31	FRACASO	11
TRABAJAR	18	SER	24	TENER	11
BAILAR	16	VALGO	20	ENFRENTAR	11
ESCRIBIR	15	ACEPTO	17	SER	10
HABLAR	14	VALORO	16	SOLA	9
REALIZAR	14	ESPECIAL	14	CARRERA	8
TRABAJO(s)	13	FORMA	14	SOLEDAD	8
ESCUCHAR	12	QUIERO	14	TRABAJO	8
TAREA(s)	12	BUENA	13	EXÁMENES	7
AMIGOS	9	INTELIGENTE	13	GENTE	7
DEMÁS	8	CAPAZ	11	TRABAJAR	7
GENTE	8	TENGO	11	VIDA	7
JUGAR	8	ÚNICO	10	FUTURO	6
REDACTAR	8	MEJOR	9	PROBLEMAS	6
PLATICAR	7	FELIZ	7	REPROBAR	6

Análisis de Correspondencias

De acuerdo con la selección de formas gráficas y segmentos repetidos se construye para cada subconjunto del corpus una tabla léxica agregada, que cruza palabras con la variable de hábitos de estudio; como ilustrativas fueron consideradas otras variables: ocupación de los padres, género, etc.

En la variable hábitos de estudio se consideran las siguientes siete áreas: Distribución de Tiempo (DT), Motivación frente al Estudio (ME), Distractores durante el Estudio (DE), Notas de Clase (NC), Optimización de la Lectura (OL), Preparación para los Exámenes (PE), Actitudes y Conductas productivas frente al estudio (AC).

Se aplica el Análisis de Correspondencias a cada una de las tres tablas léxicas que se construyen; cada tabla cruza el número de palabras a tomar en cuenta después del filtro con un total de 28 variables. Se aplica un Análisis de Correspondencia tomando en cuenta la variable Hábitos de Estudio como activa y las demás como ilustrativas. Para cada análisis se identifican las palabras y segmentos de elevada contribución y calidad de representación, estos son representados en el primer plano factorial.

Valores propios

La tabla 3 contiene los primeros dos valores propios para cada uno de los tres AC. De acuerdo con la inercia explicada por los dos primeros factores, en los tres casos los factores del primer plano explican más del 22% de la dispersión de los datos.

Tabla 3. Primeros dos valores propios para cada uno de los tres AC

	1. Soy muy bueno(a) a la hora de...		2. Me quiero a mí mismo(a) porque...		3. No me asusta...	
	F1	F2	F1	F2	F1	F2
Valor propio	0.0348	0.0199	0.0182	0.0125	0.0307	0.0278
% de inercia	17.24	9.88	13.40	9.23	12.86	11.64
% acumulado	17.24	27.12	13.40	22.63	12.86	24.50

Representación de las palabras de mayor contribución y mejor calidad de representación en el primer plano factorial

El AC identifica las palabras que contribuyen a la inercia de los diferentes ejes seleccionados. Dos conjuntos de palabras están asociados a un mismo eje, uno a la parte positiva, el otro a la parte negativa.

Una misma palabra puede estar asociada a diferentes ejes; dependiendo del contexto puede pertenecer a diferentes conjuntos. Las palabras se representan gráficamente; en cada plano factorial fueron seleccionadas las palabras más contributivas y de mejor calidad de representación. Se presenta para la primera frase un extracto de las contribuciones de las palabras a la formación de los ejes y su coseno cuadrado; se omiten los datos para las respuestas libres de las frases restantes.

1. Soy muy buena (o) a la hora de...

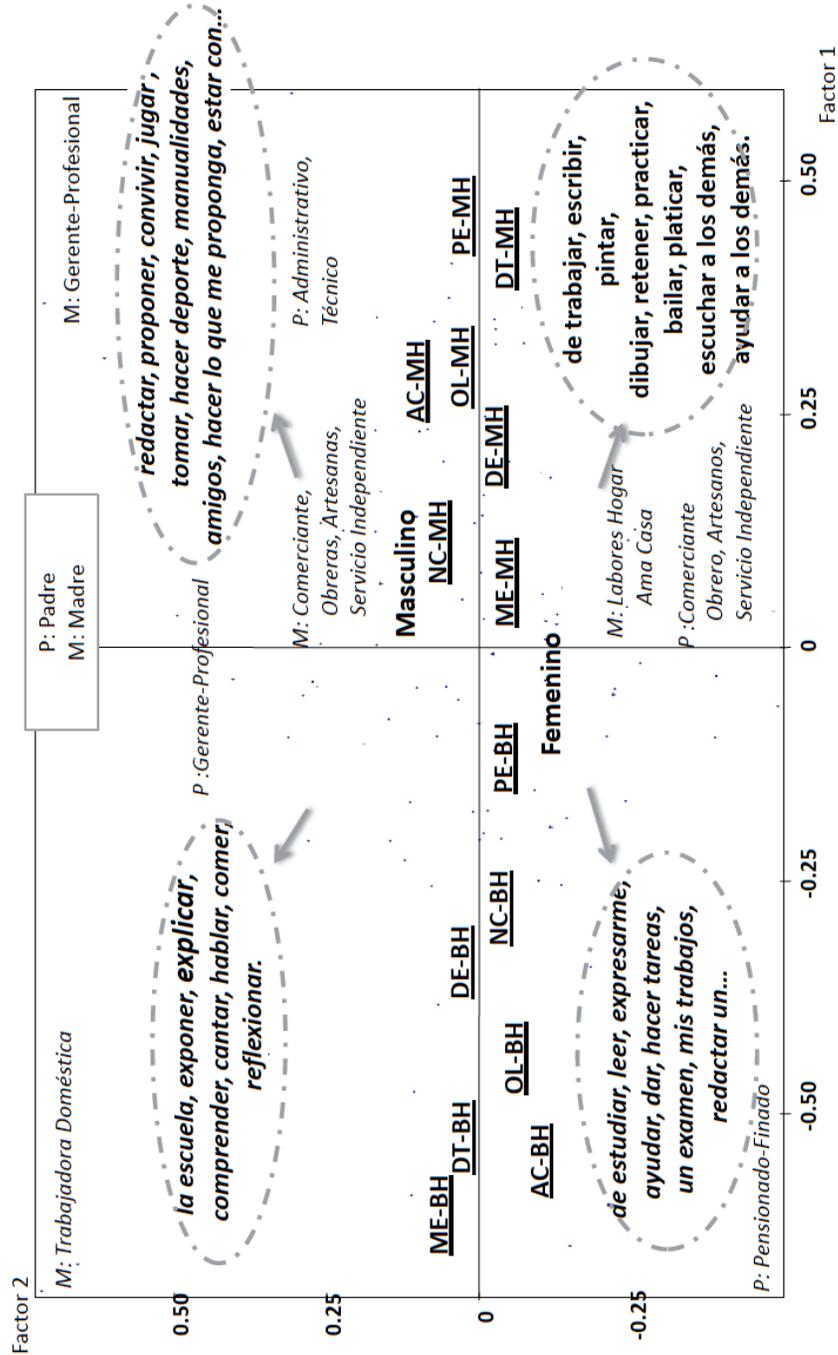
La tabla 4 presenta un extracto de las contribuciones de palabras y segmentos, y su respectiva calidad de representación.

Tabla 4. Extracto de las contribuciones de palabras y segmentos, y su respectiva calidad de representación ante la frase: Soy muy bueno(a) a la hora de...

Palabras	Contribución		Coseno cuadrado	
	F1	F2	F1	F2
Amigos	0.9	1.4	0.14	0.12
Bailar	4.1	0.1	0.37	0.01
Comer	0.8	3.4	0.06	0.15
Comprender	3.7	1.3	0.22	0.04
Convivir	2.2	5.1	0.16	0.21
Deporte	3.2	0.1	0.31	0.00
Dibujar	3.0	0.1	0.39	0.01
...	:	:	:	:
Platicar	5.0	0.0	0.39	0.00
Practicar	3.0	0.0	0.35	0.00
Tarea	6.5	2.1	0.40	0.07
Trabajo	2.0	0.3	0.20	0.02
Hacer amigos	3.6	1.6	0.29	0.07

La figura 1 representa en el primer plano factorial, las palabras de mayor contribución y mejor representadas que describen en qué se consideran muy buenos los estudiantes.

Figura 1. Representación del vocabulario utilizado por los estudiantes para decir que son buenos a la hora de..., en el primer factorial



- Los buenos hábitos aparecen próximos entre sí y en oposición a todos los malos hábitos que también aparecen próximos entre sí.
- Los estudiantes que presentan buenos hábitos de estudio dicen, entre otros, ser muy buenos a la hora de “*estudiar*”, “*leer*”, “*exponer*”, “*explicar*”, “*comprender*”, “*hacer mis trabajos y tareas*”; aunque, de acuerdo con su contribución al segundo factor, aparecen oposiciones entre “*estudiar*”, “*leer*”, “*hacer un examen*”, etc., con respuestas como “*exponer*”, “*explicar*”, “*comprender*”, “*cantar*”, “*hablar*” y “*reflexionar*”. Es importante destacar, que con respecto a las asociaciones con los hábitos de estudio, los primeros están más asociados a la optimización de la lectura y las actitudes y conductas productivas frente al estudio; los segundos se asocian principalmente con la motivación frente al estudio.
- En el otro extremo del factor 1 aparecen caracterizados los estudiantes que manifiestan tener malos hábitos en las distintas áreas; ellos dicen ser muy buenos a la hora de “*jugar*”, “*bailar*”, “*pintar*”, “*dibujar*”, “*escuchar a los demás*”, “*hacer manualidades*”, “*hacer lo que me proponga*”, entre otros. También en este caso se observa que “*trabajar*”, “*escribir*”, “*pintar*”, “*dibujar*”, “*retener*”, etc., de acuerdo con las asociaciones más fuertes, son expresiones propias de estudiantes que presentan principalmente problemas en la distribución de su tiempo; estas palabras, se encuentran en oposición a expresiones como “*convivir*”, “*tomar*”, “*jugar*”, “*hacer deporte o manualidades*”, etc., que se presentan principalmente asociadas con estudiantes que manifestaron malos hábitos en actitudes y conductas productivas frente al estudio.
- Se destacan asociaciones entre los hijos de madres y padres que tienen un puesto ejecutivo, con los estudiantes que expresan que son buenos a la hora de “*exponer*”, “*convivir*”, “*jugar*”, “*hacer deporte*”, “*hacer amigos*”, “*hacer lo que me proponga*”, etc.; con las demás ocupaciones también aparecen ciertas tendencias aunque no hay asociaciones claras.

2. Me quiero a mí mismo(a) porque...

La tabla 5 contiene las contribuciones más altas de palabras y segmentos que se utilizan para ser representados gráficamente.

Tabla 5. Extracto de las contribuciones de palabras y segmentos, y su respectiva calidad de representación ante la frase: *Me quiero a mí mismo(a) porque...*

Palabras	Contribución		Coseno cuadrado	
	F1	F2	F1	F2
Amigable	3.5	0.0	0.37	0.00
Bonita	0.7	0.8	0.11	0.08
Buena	3.9	1.6	0.49	0.14
Feliz	2.7	0.0	0.24	0.00
Honesta	4.5	1.0	0.34	0.05
Inteligente	2.3	0.2	0.28	0.01
:	:	:	:	:
Soy inteligente	1.2	0.2	0.11	0.01
Soy única	0.0	5.6	0.00	0.37
Soy único	1.5	4.2	0.10	0.45
Soy buena persona	0.0	5.4	0.00	0.34

La figura 2 muestra en el primer plano factorial las palabras que utilizaron los estudiantes al referirse a por qué se quieren a sí mismo(as), se dejan ver con claridad las asociaciones que éstas tienen con los malos y buenos hábitos de estudio y que nuevamente los buenos y malos hábitos se encuentran próximos entre sí:

- Las palabras más utilizadas por estudiantes con buenos hábitos son, por ejemplo, *“inteligente”*, *“capaz”*, *“honesta”*, y también hacen referencia a quererse porque *“soy bonita”*, *“buena persona”*, y dicen además que *“me valoro y acepto”*. En oposición sobre el primer factor, los estudiantes que tienen malos hábitos dicen quererse a sí mismo porque *“soy agradable”*, *“amigable”*, *“sociable”*, *“feliz”*, *“importante”*, etc.
- Se ven oposiciones fuertes de los estudiantes con buenos hábitos de estudio en las áreas de actitudes y conductas productivas frente al estudio y que se encuentran motivados frente al estudio, con los estudiantes que manifiestan tener malos hábitos en la forma como distribuyen su tiempo y la forma como preparan sus exámenes.
- De acuerdo con las variables ilustrativas, aparecen asociaciones de los estudiantes con buenos hábitos de estudio y calificaciones de 9 o 10 y de estudiantes con malos hábitos y calificaciones de 6 y menos de 8. Así mismo, aparecen las escuelas oficiales incorporadas al estado más asociadas con los estudiantes de buenos hábitos y las federales más asociadas con los estudiantes de malos hábitos.

3. No me asusta...

La tabla 6 muestra las palabras de contribuciones más altas y segmentos que se utilizan para ser representados gráficamente.

Tabla 6. Extracto de las contribuciones de palabras y segmentos, y su respectiva calidad de representación ante la frase: *No me asusta...*

Palabras	Contribución		Coseno cuadrado	
	F1	F2	F1	F2
Conocer	2.7	0.1	0.28	0.01
Demás	3.0	0.9	0.22	0.06
Enfrentar	0.1	2.1	0.01	0.17
Equivocarme	1.6	0.8	0.21	0.09
Errores	1.1	1.0	0.08	0.06
Estudiar	0.5	1.5	0.04	0.09
:	:	:	:	:
Sola	4.1	0.3	0.31	0.02
Terminar	3.4	3.8	0.10	0.10
Trabajo	1.1	0.0	0.10	0.00
Estar sola	4.8	0.1	0.34	0.01
Nuevos retos	2.2	0.1	0.23	0.01

En la figura 3 se observan en el primer plano factorial las frases y segmentos repetidos de mayor contribución y mejor representadas por algunos estudiantes que expresan que no les asusta “*estar sola*”, “*problemas*”, “*carrera*”, “*escuela*”, “*obstáculos*”, “*equivocarse*”, “*reprobar*”. Estas expresiones muestran a estudiantes que tienen buenos hábitos, principalmente del género femenino, y presentan promedio general de bachillerato por arriba del 8 de calificación; en oposición se encuentran los estudiantes que muestran malos hábitos y en su gran mayoría son del género masculino; estos dicen que no les asusta “*fracasar*”, “*estudiar*”, “*retos*”, “*oscuridad*”, “*errores*”, “*perder*”, “*maestros*”, “*morir*”, “*soledad*”, “*enfrentar*”, “*nuevos retos*”, presentan promedio general de bachillerato por debajo del 8 de calificación.

4. *Mi rendimiento escolar lo considero ...*

Para las respuestas de esta frase incompleta se presentan las estadísticas preliminares a cualquier otro análisis factorial multivariado que se deseara hacer, suficientes para cubrir los objetivos de este trabajo (tablas 7 y 8).

Tabla 7. Vocabulario utilizado ante la frase: *Mi rendimiento escolar lo considero...*

Palabras repetidas	longitud	frecuencia
Adecuado	8	3
Bien	4	4
Bueno	5	268
Eficiente	9	3
Excelente	17	12
Medio	5	6
Muy	3	35
Regular	7	51

Tabla 8. Resumen del vocabulario utilizado ante la frase: *Mi rendimiento escolar lo considero...*

Palabras repetidas	frecuencia	%
Excelente, muy bueno	44	12
Adecuado, bien, bueno, eficiente	278	73
Medio, regular	57	15
Total	379	100

De acuerdo con las respuestas que los estudiantes dan sobre su rendimiento escolar, se puede resumir que la mayoría (el 73%) considera que su rendimiento es adecuado, eficiente, bueno o bien. El 12% dice tener un excelente o muy buen rendimiento y el 15% reconoce que es regular o de rendimiento medio.

Conclusiones

La riqueza de información que se tiene con esta primera etapa del proyecto es de utilidad a los tutores académicos de la Licenciatura en Pedagogía porque tienen a su alcance resultados de evaluación de condiciones de estudios de cada uno de sus tutorados; adicionalmente cuentan con los reportes que se les proporcionaron cuyo contenido son recomendaciones específicas para mejorar el rendimiento escolar de los estudiantes, su postura frente al Modelo Educativo Integral Flexible (MEIF), los tiempos diversos que plantean para su proyecto académico y que no necesariamente han establecido conjuntamente con su tutor (siendo que es una de las tareas esenciales en su función de tutores), el impacto de la actividad de los padres en el desempeño académico de sus hijos, entre otras.

Los métodos y técnicas empleados para el análisis de datos de naturaleza textual, son un instrumento muy útil para el análisis de frases incompletas con respuestas libres, por lo que el Análisis de Correspondencia ha permitido detectar semejanzas y oposiciones entre estudiantes con respecto al vocabulario empleado en sus respuestas y asociaciones significativas con los hábitos de estudio que manifestaron, así como con algunas de las variables cerradas contenidas en la encuesta.

En la frase “Soy bueno (a) a la hora de...” encontramos 3 grupos de estudiantes: los que tienen buenos hábitos de estudio y que por ende expresan ser buenos para *estudiar, leer, realizar sus trabajos o tareas, exponer, explicar, y/o comprender*, son personas cuyo éxito radica principalmente en que dedican tiempo al estudio. Los otros dos grupos que presentan hábitos inadecuados se diferencian en que un grupo expresa no dedicarle mucho tiempo al estudio porque le gusta realizar otras actividades (por ejemplo: *pintar, practicar, bailar*); el otro grupo, está representado por jóvenes con padres que tienen puestos ejecutivos altos. Esta reflexión nos lleva a preguntarnos si pudiera ser que los padres profesionistas y/o con un alto puesto de trabajo, están descuidando el rendimiento escolar de sus hijos. En todo caso, de acuerdo con estos resultados es determinante que *las Actitudes y Conductas productivas frente al estudio, la Optimización de la Lectura y la Distribución de Tiempo* son áreas indispensables que los estudiantes deben fortalecer para mejorar su rendimiento escolar.

Ante la frase “Me quiero a mí mismo (a) porque...” encontramos que los buenos hábitos de estudio son propios del estudiante que manifiesta valoración en sí mismo expresada como: *soy honesta, capaz, buena persona*, expresiones que reflejan fortaleza en su propia su axiología. Mientras, los estudiantes con malos hábitos de estudio dejan ver que su valor radica en ser *amigable, original, o feliz*.

De la pregunta “No me asusta...” se desprende que los estudiantes que dicen no asustarse por *problemas, ni por la escuela, obstáculos, por equivocarse o reprobado*, son personas con buenos hábitos de estudio; esto probablemente es lo que permite que enfrenten sus estudios con decisión y entereza. Los de malos hábitos se refieren a situaciones cotidianas que la vida nos presenta como *errores, perder o enfrentar con éxito nuevos retos*.

En relación a la frase “Mi rendimiento escolar lo considero...” la gran mayoría de los estudiantes se visualiza con un rendimiento escolar adecuado, eficiente, bien o bueno. Lo rescatable para una toma de decisiones con respecto a mejorar dichas condiciones de los alumnos-tutorados de Pedagogía es que un mínimo porcentaje refiere que su rendimiento es muy bueno o excelente.

Por último, el presente trabajo le brinda información a la coordinación del Departamento de Orientación Educativa de la Facultad de Pedagogía para llevar un seguimiento puntual ante la falta de vocación por la profesión, la presencia de hábitos de estudio inadecuados lo que ofrece un área de oportunidad para intervenir este departamento, además de orientar sobre las áreas terminales y campo de trabajo del futuro pedagogo.

A partir de este trabajo se pueden derivar ideas para nuevas investigaciones que amplíen el conocimiento de los estudiantes y que sirvan para mejorar su rendimiento escolar.

Referencias

- Abascal, E. y Franco, M. (2002). *Análisis textual de encuestas: aplicación al estudio de las motivaciones de los estudiantes en la elección de su titulación*. Metodología de Encuestas, vol. 4, n° 2, 2002, 195-209.
- Bécue, M. (1991). Análisis de datos textuales. París: CISIA.
- Benzécri, J. (1976). L'Analyse des Données I. La Taxonomie. L'Analyse des Données II. L'Analyse des Correspondences. París: Dunod.
- Escofier, B. y Cordier, B. (1965). L'analyse des correspondances. Thèse, Faculté des Sciences de Rennes ; publiée en 1969 dans les Chiers du Bureau Universitaire de Recherche Opérationnelle, n° 13.
- Guttman, L. (1941). The quantification of a class of attributes: a theory and method of a scale construction. In: The prediction of personal adjustment (Horst P., ed.) New York: SSCR, 251-264.
- Hayashi, C. (1956). Theory and examples of quantification. (II) Proc. Of the

Institute of Statist. Math. 4 (2), 19–30.

Hernández, M. (2007). “Métodos para el Análisis Factorial de una tabla de contingencia múltiple aplicación a la comparación de dos corpus cronológicos” Universidad Politécnica de Cataluña, Dpto. EIO.

Lebart, L. y Salem, A. (1989). *Analyse Statistique des Données Textuelles*. Paris: Dunod.

Lebart, L., Morineau, A. y Bécue, M., (1989). *SPAD-T Système Portable pour l’Analyse des Données Textuelles. Manuel de l’utilisateur*. Paris: CISIA.

Lebart, L., Salem A. y Berry, E. (1998). *Exploring Textual Data*, Dordrecht: Kluwer.

Lebart, L., Morineau A. y Piron, M. (2000). *Statistique exploratoire multidimensionnelle*. Paris: Dunod.