

ISSN – 0325 – 6308

CUADERNOS UNIVERSITARIOS

Universidad Nacional del Comahue

Centro Regional Universitario Bariloche



Nº 56 – Febrero de 2012

ASESORAMIENTO Y CAPACITACIÓN A DISTANCIA PARA INVESTIGADORES EN
CIENCIAS SOCIALES EN EL USO DE ANÁLISIS MULTIVARIADOS

Volumen 1: La lexicometría

Montserrat de la Cruz, Nora Scheuer y Ana Pedrazzini

Índice

1. Presentación de la modalidad de asesoramiento	3
Equipo a cargo	3
Asesoramientos efectivizados.....	4
¿Por qué asesorar y capacitar a los investigadores en ciencias sociales en el uso de métodos de análisis multivariado?.....	4
¿Qué buscamos con este asesoramiento y capacitación?	5
¿Cómo trabajar durante el asesoramiento?.....	5
2. Una presentación de los métodos	6
¿Qué es la lexicometría o análisis de datos textuales?	6
¿Qué es el Análisis Factorial de Correspondencias aplicado a la categorización de la información?.....	8
3. Introducción al método lexicométrico	12
Ventajas y desventajas	13
4. Información sobre el estudio mediante el cual ilustramos conceptos y procedimientos lexicométricos básicos	14
5. Conceptos lexicométricos básicos.....	16
6. Pasos para implementar la lexicometría con el programa SPAD.T.....	22
Carga de los ficheros: cómo ingresar los datos	22
Instrucciones para activar los procesamientos.....	26
7. Las salidas de los procesamientos en el estudio utilizado como ejemplo.....	36
Primera etapa: poner a prueba de la hipótesis de diferenciación léxica	36
Segunda etapa: identificar grupos léxicos	41
Tercera etapa: describir respuestas modales	44
Apéndice: Algunas aclaraciones	48
Agradecimientos	51
Referencias	51

1. Presentación de la modalidad de asesoramiento

A partir de numerosas consultas recibidas desde hace más de una década por parte de equipos de investigación y estudiantes de maestría y doctorado de universidades de Argentina, España y Estados Unidos, decidimos implementar una modalidad de asesoramiento y capacitación a distancia dirigida a investigadores en ciencias sociales, mediante el uso de las nuevas tecnologías de la información y la comunicación (correo electrónico, Skype). Nos proponemos principalmente favorecer el conocimiento de:

- la implementación de tres métodos estadísticos de análisis multivariados, el Análisis de datos textuales o lexicometría (aplicado al análisis de corpus), el Análisis Factorial de Correspondencias (aplicado sea a datos textuales o a categorizaciones de información primaria o secundaria) y el Análisis Factorial de Correspondencias Múltiples (aplicado a respuestas a cuestionarios con preguntas cerradas de opción múltiple y a categorizaciones de información primaria o secundaria), con el programa SPAD (Decisia, 2002).
- la delimitación y operativización de las variables relevantes estudiadas (de los individuos, de la tarea, de la situación);
- la preparación de los materiales recogidos para su tratamiento de acuerdo a estos métodos;
- los procedimientos comprendidos en el o los métodos seleccionados para el análisis de dichos materiales;
- la toma de decisiones a partir de los productos obtenidos con cada uno de los procedimientos utilizados;
- la interpretación de los resultados obtenidos a partir de la aplicación de cada uno de los procedimientos y
- la elaboración de informes de investigación sobre los resultados obtenidos y la metodología utilizada.

Equipo a cargo

El equipo a cargo de la asesoría y capacitación a distancia está integrado por:

Dra. Nora Scheuer. Investigadora Independiente del Consejo Nacional de Investigaciones Científicas y Tecnológicas (CONICET) en Centro Regional Universitario Bariloche (CRUB) de la Universidad Nacional del Comahue (UNCo).
Área: Psicología.

Mg. Montserrat de la Cruz. Profesora Titular, CRUB, UNCo. Área: Psicología.

Dra. Ana Pedrazzini. Becaria postdoctoral de CONICET en CRUB, UNCo. Área: Ciencias de la Información y de la Comunicación.

Asesoramientos efectivizados

Hasta la fecha hemos asesorado en el uso de los métodos estadísticos mencionados a estudiantes e investigadores de las siguientes universidades:

Universidad de Buenos Aires (Facultad de Psicología y Facultad de Ciencias Sociales)

Universidad Nacional de Río Cuarto (Facultad de Ciencias de la Educación)

Universidad Autónoma de Madrid (Departamento de Psicología Básica)

Universidad Complutense de Madrid (Facultad de Psicología)

Universidad de Barcelona (Facultad de Psicología)

Universidad de Sevilla (Facultad de Psicología)

Universidad Pablo Olavide (Facultad de Psicología)

Tufts University (Department of Education)

Université Paris IV-Sorbonne (CELSA)

¿Por qué asesorar y capacitar a los investigadores en ciencias sociales en el uso de métodos de análisis multivariado?

Porque el uso de procedimientos estadísticos potencia, enriquece y aporta mayor confiabilidad a los resultados de una investigación cualitativa. La utilización de los métodos de Análisis multivariado permite el estudio de masas voluminosas de información mediante la confrontación de forma simultánea de dos o más variables de análisis y la identificación de casos típicos. De esta manera, estos métodos aportan información acerca de las relaciones entre variables e individuos, que normalmente no podrían ser alcanzados mediante esfuerzos de sistematización y selección manuales.

¿Qué buscamos con este asesoramiento y capacitación?

El objetivo general del programa de asesoramiento y capacitación de investigadores en el área de las ciencias sociales es el desarrollo de recursos y habilidades para abordar el análisis estadístico de datos cualitativos. El programa está especialmente orientado para investigadores en formación. Los objetivos específicos apuntan a que los investigadores desarrollen la capacidad para:

- Seleccionar métodos de análisis estadísticos adecuados para el tratamiento de datos cualitativos, según el enfoque de investigación en el que se enmarquen: sus fundamentos, marco teórico, objetivos, metodología de obtención de la información, participantes.
- Aplicar dichos análisis a un estudio concreto y acotado.
- Interpretar los resultados.
- Informar el método estadístico utilizado y los resultados obtenidos.
- Utilizar estos métodos en un nuevo estudio en forma más autónoma.

¿Cómo trabajar durante el asesoramiento?

A partir de un acuerdo entre el investigador interesado (I) y el equipo a cargo del asesoramiento (EA), se pautará un cronograma cuyas etapas principales son:

1. Lectura de este texto y de otros documentos sugeridos.
2. Formulación de objetivos del estudio de aplicación (en general es un sub-estudio del estudio más amplio en curso; por ejemplo, tres preguntas en una entrevista, o las respuestas a una de las partes de un cuestionario), delimitación de materiales a ser analizados, identificación de variables.
3. Preparación de las bases a ser procesadas.
4. Instalación del programa por parte del investigador en su computadora.
5. Procesamientos.
6. Redacción del informe.

2. Una presentación de los métodos

A continuación presentamos brevemente los dos métodos objeto del asesoramiento, para en un segundo paso ejemplificar su aplicación. Nos ocuparemos de:

- El análisis de datos textuales, o lexicometría, aplicado a textos diversos e implementado mediante algunos de los procedimientos que figuran entre los métodos “Analyses Textuelles” y “Analyses Factorielles” del programa SPAD (Decisia, 2002). El presente volumen se dedica especialmente al método lexicométrico.
- El Análisis Factorial de Correspondencias y el Análisis Factorial de Correspondencias Múltiples, aplicados a análisis de conjuntos de preguntas cerradas en cuestionarios y a la categorización de la información, e implementados mediante algunos de los procedimientos que figuran entre los métodos “Analyses Factorielles” del programa SPAD (Decisia, 2002). El siguiente volumen de este trabajo aborda estos dos métodos en mayor profundidad.

En la breve presentación que figura a continuación, pretendemos brindar un panorama de conjunto de lo que cada uno de estos métodos aporta. En las aplicaciones de cada uno de los métodos que desarrollamos en este volumen y en el siguiente respectivamente explicamos el sentido de los mismos y detallamos sus pasos. Para una comprensión más completa remitimos a la bibliografía específica.

¿Qué es la lexicometría o análisis de datos textuales?

Los enfoques lexicométricos o de la estadística textual, que se apoyan en las técnicas estadísticas desarrolladas por la Escuela Francesa de Análisis de Datos (Lébart, Salem, Bécue y Bertaut, 2000), toman como unidad de análisis básica la palabra. Resultan muy útiles para el tratamiento de corpus de diverso origen y características. Entre ellos:

- la transcripción de entrevistas;
- las respuestas a preguntas abiertas en cuestionarios;
- la transcripción de las expresiones orales de participantes en situaciones grupales

como la sala de clases, sesiones de terapia, etc.;

- las producciones escritas en entornos virtuales; textos literarios, periodísticos, científicos, etc.; discursos políticos.

Mediante la aplicación inicial de un Análisis Factorial de Correspondencias (AFC) (Crivisqui, 1993), la lexicometría estudia en un primer paso las asociaciones entre los “individuos” (participantes, autores, textos, etc.) y todas las palabras diferentes que han enunciado con cierta repetitividad (delimitada por un umbral de frecuencia), sin ningún tipo de manipulación del corpus. Este AFC permite identificar grupos léxicos, caracterizados por algunas palabras y algunos individuos, e ilustrados por modalidades de variables de los individuos (tales como grado escolar, sector social, género, etc., si los individuos son personas) o de la situación (tiempo, etc., si los individuos son producciones). De acuerdo a las asociaciones encontradas -que se pueden visualizar en planos factoriales-, el investigador puede optar por agrupar los individuos según la o las variables que hayan resultado relevantes (área de conocimiento, grado escolar, sector social, género, período temporal, etc.) y, en algunas ocasiones -como veremos más adelante, no siempre es necesario-, a estudiar mediante un nuevo AFC las asociaciones entre palabras y cada texto colectivo conformado según las variables consideradas (por ejemplo, para la variable grado escolar, cada texto colectivo reúne todas las respuestas dadas en una entrevista por los niños de los diversos grados indagados). Este segundo AFC también posibilita identificar *grupos léxicos*, los que en este caso estarán caracterizados por el uso de algunas palabras y, por lo general, también por alguna modalidad de la o las variables consideradas. Por ejemplo, permite delimitar las palabras que caracterizan a las producciones lingüísticas de niños de distintos grados escolares, sector sociocultural, género, etc.

Un procedimiento complementario (llamado: Selección Automática de Respuestas Modales) ordena de acuerdo a su tipicidad todas las respuestas (o textos individuales) que componen cada texto colectivo. Este ordenamiento está basado en la distancia χ^2 . La respuesta que aparece en primer lugar es la que representa en forma más completa el texto colectivo y luego sigue un orden decreciente. Este procedimiento es especialmente útil porque permite recuperar las respuestas originales completas proporcionadas por los individuos e identificar aquellos que resultan emblemáticos de cada modalidad de las variables asociadas a determinados grupos léxicos, para así llevar

a cabo la descripción de los grupos. Generalmente se consideran varias respuestas modales para cada modalidad, ya que una sola no alcanza a condensar las características del texto colectivo. Asimismo, este procedimiento ofrece criterios más objetivos para la selección de textos (sean respuestas orales o producciones escritas) ilustrativos que los habitualmente utilizados en la investigación en ciencias sociales. Esto se debe a que los textos utilizados para la descripción de grupos de individuos no son elegidos por los investigadores, sino delimitados y validados por un conjunto de procedimientos estadísticos. Tales textos posibilitan aproximarse a modos *singulares* de producción en un contexto dado, a la vez que aportan una “especie de resumen de los datos iniciales, que se obtuvo sin ninguna codificación ni interpretación previa del texto” (Lebart, Salem y Bécue-Bertaut, 2000, pág. 183), que se encuentra validado de las formas de expresarse de una *categoría* de individuos.

En conjunto, el método lexicométrico posibilita una lectura de los textos sustancialmente distinta y complementaria a las realizadas desde un enfoque cualitativo tradicional. Permite acceder a la información de modo confiable, profundizar comprensiones de procesos complejos (como el aprendizaje, los cambios discursivos a partir de una intervención educativa o terapéutica, el posicionamiento enunciativo de un autor, etc.), sintetizarlos e ilustrarlos, como dijimos, con ejemplos típicos (respuestas modales), contribuyendo a recuperar las respuestas “con toda su originalidad y, a veces, su carga emotiva” (Bécue, 1991). Mantiene al investigador cerca de lo expresado por los individuos investigados, le permite hacer explícitas sus decisiones y fundamentarlas desde los textos originales, así como también los resultados de los procedimientos y la teoría en un proceso espiralado y progresivo.

Una explicación más completa de los fundamentos y procedimientos de la aplicación de la lexicometría a la investigación por medio de entrevistas de las concepciones de niños, jóvenes y adultos en el dominio psicológico puede consultarse en Baccalá, de la Cruz y Scheuer (2002).

¿Qué es el Análisis Factorial de Correspondencias aplicado a la categorización de la información?

Otro modo de estudiar las producciones de los sujetos investigados se basa en la delimitación de dimensiones y categorías de análisis, con la consecuente codificación del material en estas categorías. En este caso, los procesamientos estadísticos se

aplicarán sobre los resultados derivados de esas codificaciones. La unidad de análisis ya no es la palabra, sino aquello que es objeto de categorización (si se categoriza el foco predominante de la respuesta a una pregunta, la unidad de análisis es la respuesta completa de cada individuo; si se categoriza la dirección de la mirada al realizar algo, la unidad es cada fijación de la mirada en una dirección). En otras palabras, en lugar de procesar en forma casi directa las producciones de los sujetos, el objeto del procesamiento es la codificación implementada por los investigadores sobre las producciones, expresiones o comportamientos (verbales o no verbales) de los sujetos. ¿Qué comparte entonces esta aproximación categorial al tratamiento de la información con la lexicometría? Comparte el enfoque estadístico, orientado al estudio confiable de las asociaciones entre los individuos y las diversas variables al interior del universo estudiado, sin la pretensión de generalizar las descripciones alcanzadas a otros contextos.

Cuando este método se aplica a los mismos tipos de corpus indicados para el método lexicométrico (transcripciones de entrevistas; respuestas a preguntas abiertas en cuestionarios; transcripciones de las expresiones orales de participantes en situaciones grupales como la sala de clases, sesiones de terapia, etc.; producciones escritas en entornos virtuales; textos literarios, periodísticos, científicos, etc.; discursos políticos), la diferencia radica en el tipo de tratamiento implementado. En el caso del análisis categorial se pone en juego un abordaje a la vez más mediado por los investigadores y más distanciado de las producciones originales de los sujetos.

Como hemos indicado antes, los procesos de categorización de las producciones de los sujetos para su posterior tratamiento estadístico mediante análisis factoriales de correspondencias pueden ser aplicados, además, a otros tipos de materiales. Entre ellos:

- las respuestas a preguntas cerradas de elección múltiple en cuestionarios;
- dimensiones no específicamente lingüísticas de las producciones verbales: aspectos plásticos y paratextuales de las producciones escritas; aspectos prosódicos de las producciones orales; aspectos pragmáticos como las intenciones, alternancia y efectos de lo dicho; etc;
- producciones gráficas no verbales;
- expresiones gestuales;

- otras conductas corporales.

En todos estos distintos casos, el proceso de categorización de la información comienza por la delimitación de dimensiones de análisis (a qué aspectos de la información interesa atender) y, al interior de estas dimensiones, la delimitación de categorías específicas que distinguen entre diferentes modos en que la dimensión en cuestión se manifiesta en el material analizado. El número de categorías en las distintas dimensiones puede variar de acuerdo a la dimensión. La delimitación, tanto de las dimensiones como de las categorías, puede recuperar en forma fiel o adaptada las generadas en otro estudio empírico o teórico, combinar las definidas en varios otros estudios antecedentes, o ser el producto original del estudio en curso. En cualquiera de los casos, el proceso de categorización comprende tareas de:

- delimitación: cuáles son las dimensiones y las categorías que se tienen en cuenta -que nunca agotan el universo de lo posible, sino que se seleccionan según los focos y posibilidades de la investigación- y cuál es la unidad de análisis que se considera para cada dimensión (por ejemplo, la respuesta completa de cada participante a una de las muchas preguntas de una entrevista; cada turno en una conversación; el texto producido en una misma fecha en un diario de clase; etc.);
- definición: qué es cada dimensión y cada categoría, según su comprensión y/o extensión;
- operativización: cómo se reconocen en el material analizado las diversas dimensiones y categorías;
- codificación: para cada dimensión de análisis, a qué categoría corresponde cada ocurrencia de la unidad de análisis;
- control inter-investigador: en qué medida coinciden las codificaciones de las mismas ocurrencias realizadas por distintos investigadores.

Una vez que este proceso -que como veremos más adelante dista de ser lineal- esté completado y sus resultados se encuentren volcados en tablas adecuadas según las distribuciones de frecuencias resultantes, se evalúa la factibilidad de aplicar uno o más Análisis Factoriales de Correspondencias:

- **Simples (AFC)**, para estudiar las asociaciones entre individuos o grupos de individuos (por ejemplo, todos los alumnos de una misma escuela o de un mismo grado escolar, todos los autores de un mismo periódico o de un período temporal de publicación de ese periódico, etc.) por una parte, y todas las categorías relativas a una o más dimensiones de análisis, por otra. Es decir, el AFC cruza las modalidades de dos variables categóricas (individuos o grupos x palabras), una de las cuales desplegada en las columnas y la otra en las filas de una tabla de contingencia. El AFC nos brinda una comparación (semejanza o diferencia) entre perfiles de distribución de las filas en las columnas y de los perfiles de distribución de las columnas en las filas. Dos perfiles fila serán más parecidos cuando su distribución en las columnas sea más similar y contribuirán más a la inercia o variabilidad total cuanto más alejado del perfil medio se encuentren. Lo mismo se aplica a los perfiles columna. El AFC también nos brinda un interpretación en términos de asociaciones de filas con columnas, en términos baricéntricos. Es decir una columna es el baricentro (salvo una dilatación, pues está multiplicado por un número distinto de uno) de las filas cuando se las pondera por su peso en dicha columna.
- **Múltiples (AFCM)**, para estudiar las asociaciones entre individuos y las diversas dimensiones de análisis según las categorías que los individuos adoptan para cada dimensión. Es decir, el AFCM cruza la tabla de individuos con las modalidades de más de dos variables categóricas. El AFCM nos permite hallar grupos de individuos similares (es decir que prefieren casi las mismas modalidades de las variables) y entre modalidades. Dos modalidades serán parecidas cuando son casi los mismos individuos las que las prefieren. Por otra parte, una modalidad será el baricentro de los individuos que la toman.

Un aporte muy útil de los AFC y AFCM es que sus resultados pueden visualizarse en planos factoriales (cada uno formado por dos ejes factoriales ortogonales). De acuerdo a las asociaciones encontradas entre modalidades se procede a identificar grupos de individuos y categorías de la o las dimensiones consideradas. Asimismo, es posible identificar aquellos individuos que se encuentran más próximos al centro de gravedad de cada grupo, lo que ofrece criterios más objetivos para la selección

de ejemplos ilustrativos que los habitualmente utilizados en la investigación en ciencias sociales. De este modo, los ejemplos y las respuestas utilizadas para la descripción de grupos de individuos no son elegidas por los investigadores, sino delimitadas y validadas por procedimientos estadísticos.

La aplicación de análisis factoriales de correspondencias al término del proceso de categorización posibilita una lectura relacional de los resultados, difícilmente alcanzable sin el recurso de un instrumento estadístico. Podríamos decir que permite pasar de la descripción de repertorios de modalidades aisladas a la identificación y consecuente descripción e ilustración de relaciones entre modalidades.

Una explicación más completa de los fundamentos y procedimientos de la aplicación de análisis factoriales de correspondencias puede consultarse en las tesis de Montoro (2003) y de Torres Curth (2005), y en el Cuaderno Universitario elaborado por Baccalá y Montoro (2008).

3. Introducción al método lexicométrico

Se llama lexicometría al estudio de textos con métodos estadísticos. La lexicometría se aplica actualmente a todo tipo de textos transcritos sobre o directamente accesibles en soporte informático. Esto permite que sea utilizada entre otros métodos de aproximación al análisis de los textos (análisis del discurso; análisis de contenido) en distintas disciplinas (lingüística, historia, sociología, literatura, política, educación, psicología, psiquiatría, entre otras), teniendo en cuenta en cada caso la perspectiva propia de la investigación.

La lexicometría nace de las posibilidades de cálculo que ofrece la informática. Este tipo de técnica fue cobrando auge, tanto desde el punto de vista de su aplicación como de su desarrollo metodológico (Germano, 2002) a fines de los años 80 del siglo pasado, con el desarrollo de la micro-informática que permitió a los investigadores un acceso a una información descentralizada e interactiva (Crivisqui, 1993). Actualmente se considera que la lexicometría ha probado, en lo que concierne a los estudios textuales (Bécue, 1991; Benzécri, 1973), su aptitud para elaborar tipologías o categorías mediante el recuento de formas gráficas o palabras.

Ventajas y desventajas

Distintos autores consideran que el objeto al que se aplica el método lexicométrico es fundamentalmente socio-psicológico, ya que a través del procesamiento de las producciones lingüísticas se pretende acceder a las actitudes, representaciones, concepciones de grupos humanos (García Calvo, 1991). Se parte del supuesto de que dichas actitudes, representaciones o concepciones se manifiestan en el manejo del vocabulario. Es decir, en la elección predominante de los términos, en la frecuencia relativa de uso de unos sobre otros, en la frecuencia de las combinaciones de dos o más de ellos y en la relación de todo ello con las características de los individuos (sean entrevistados, participantes en una situación grupal, autores, etc.) según variables seleccionadas. Como todo estudio estadístico, en éste también es claro que a medida que aumenta el número de sujetos investigados, o individuos, aumenta la precisión de los resultados aportados y su alcance.

Entre las ventajas principales que suelen señalarse es que en cierto modo la aplicación del método contribuye a:

- corregir lo que hay de subjetivo y no intencionado, lo que puede estar condicionado por la propia perspectiva personal incluso en el observador más agudo y desinteresado, al precisar la presencia y el sentido de las variaciones, empezando por la determinación de lo que es azaroso y lo que debe tomarse como intencionado (sea la intención propiamente consciente y deliberada, o más bien motivada por el ámbito social, las tradiciones asumidas implícitamente por el investigador, o el contexto inmediato de producción);
- precisar las impresiones a “ojo”, las cuentas o estimaciones “por encima”, que se muestran demasiado vagas y muchas veces resultan distorsionadas;
- acceder a las transiciones léxicas en los estudios evolutivos;
- comparar universos léxicos que por su extensión difícilmente podrían ser comparados;
- mostrar tendencias que pueden orientar la elaboración de categorías de análisis;
- seleccionar mediante métodos probabilísticos textos relevantes para ilustrar las categorías.

Una de las desventajas más importantes que se señala a la lexicometría es que el

análisis de las producciones lingüísticas (habladas o escritas en el material original), se lleva a cabo sobre los elementos que suelen ser los más visibles y aparentemente los más próximos a la conciencia, tales como las palabras con significado, es decir, ciertos fragmentos o piezas del vocabulario. En cambio, los elementos menos accesibles a la conciencia, como las reglas y las estructuras de la sintaxis, las tácticas de organización pragmática del discurso, las presuposiciones más o menos reguladas gramaticalmente, presentan dificultades enormes para su programación y tratamiento y por lo tanto tienden a no ser objeto de este tipo de estudio (Bécue, 1991).

También se ha dicho por ejemplo, que los resultados del cómputo son triviales, es decir, que son los que uno podría haber visto a ojo y sin necesidad de estos procesamientos. Creemos que no siempre es así. Más aún, nos inclinamos a pensar que la profundidad de los hallazgos, como sucede con la utilización de otros métodos de análisis, depende fundamentalmente de la capacidad y agudeza del investigador.

Para adentrarnos en los conceptos lexicométricos básicos así como en los pasos de aplicación de este método, hemos optado por referirlos a un ejemplo. Por ello, antes de introducir esos contenidos, presentamos las características básicas del estudio que usamos como ejemplo.

4. Información sobre el estudio mediante el cual ilustramos conceptos y procedimientos lexicométricos básicos

En el análisis de las respuestas verbales en las entrevistas realizadas en el marco del proyecto *Las concepciones de alumnos, maestros y padres sobre el aprendizaje* (financiado por la Secretaría de Investigación de la Universidad Nacional del Comahue; 2000-2003), utilizamos el método lexicométrico o análisis estadístico de datos textuales. Dicho proyecto, como lo indica su nombre, se orientó a describir las concepciones de distintos sujetos sobre el aprendizaje de la escritura en los inicios de la escolaridad. En particular, procuró describir qué componentes del sistema de aprendizaje (condiciones, procesos y resultados) aprendices y enseñantes tienen en cuenta y de qué manera lo hacen, así como analizar los cambios o variaciones entre las concepciones, en términos de teorías implícitas. Se estudió la incidencia del grado escolar/edad en las concepciones de aprendizaje de los alumnos (preescolar, primer grado, cuarto grado),

del entorno social en las concepciones de alumnos, padres y docentes (en escuelas cuya población predominante es de sectores medios o marginados) y del género (femenino, masculino) en las concepciones de los alumnos (Scheuer, de la Cruz y Pozo, 2010).

Para desarrollar los modos de aplicación de la lexicometría nos basamos en uno de los estudios de dicho proyecto, que investigó las ideas sobre el contenido del pensamiento en distintos momentos de la actividad de escribir, expresadas por 60 niños de preescolar, primero y cuarto grado del nivel primario, que provienen de distintos entornos socioculturales (medios y marginados). El relevamiento de la información fue llevado a cabo mediante una entrevista individual sobre la base de cuatro preguntas (Scheuer, de la Cruz, Pozo, Huarte y Sola, 2006) que refieren una sucesión de cuatro momentos en los que un niño o niña (según el sexo del entrevistado) produce un texto escrito, representados mediante tarjetas gráficas (ver Figura 1).

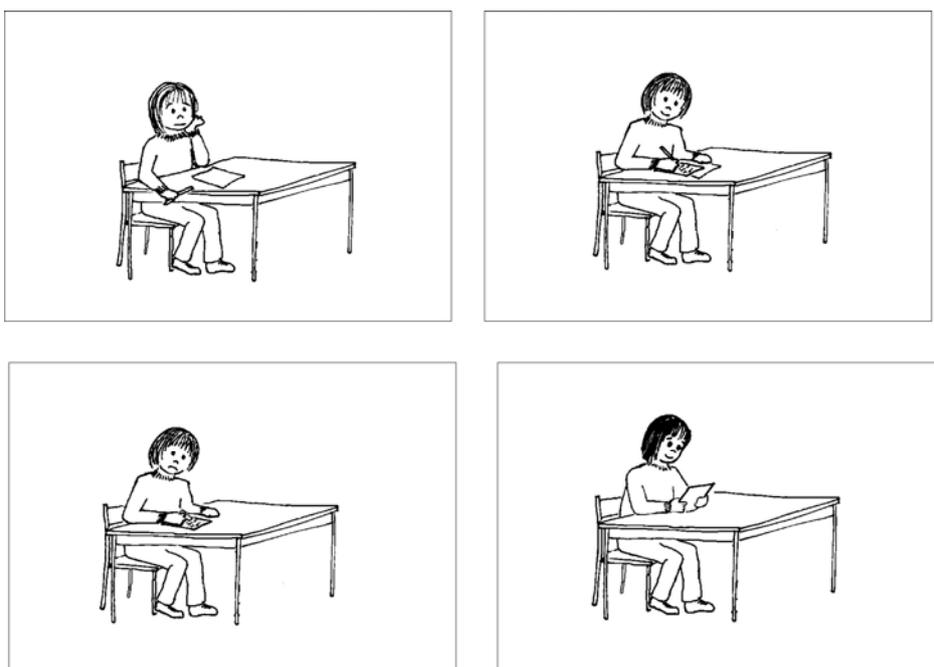


Figura 1: Tarjetas gráficas empleadas en la tarea

Las preguntas son:

Ahora el nene/a está por escribir un cuento, o una carta a los parientes que viven lejos.

El nene está por empezar a escribir. ¿Qué está pensando? (tarjeta superior izquierda).

El nene está escribiendo. ¿Qué está pensando? (tarjeta superior derecha).

El nene está borrando algo de lo que escribió. ¿Por qué será que borra? ¿Cómo se dio cuenta que hacía falta borrar? (tarjeta inferior izquierda).

El nene terminó de escribir y ahora lo está leyendo. ¿En qué estará pensando? ¿En qué se fija? (tarjeta inferior derecha).

Los resultados del análisis de estas cuatro preguntas fueron sistematizados en diversos trabajos. Según una perspectiva metodológica, en: Baccalá, de la Cruz y Scheuer (2002). Desde una perspectiva evolutivo/educativa, en: Scheuer, de la Cruz, Pozo, Huarte y Sola (2006). También estudiamos las respuestas verbales de maestros y padres mediante la aplicación de la lexicometría, tal como se informó en: de la Cruz, Scheuer, Huarte, Sola (2002).

5. Conceptos lexicométricos básicos

El corpus. Segmentación. El corpus es una sucesión de ocurrencias de formas gráficas y de delimitadores. Se segmenta en unidades mínimas como las formas gráficas o palabras, las que se definen por la sucesión de caracteres comprendidos entre espacios y delimitadores. Una forma gráfica o palabra se caracteriza por su frecuencia en el corpus. El corpus se caracteriza por la frecuencia máxima y mínima de cada una de las palabras distintas que lo componen, así como por la distribución de las palabras según su frecuencia. Una palabra utilizada una sola vez se llama “hapax”.

Unidades de la estadística léxica. La unidad adoptada generalmente en los análisis automáticos es la palabra, definida como la sucesión de caracteres no delimitadores (normalmente letras) junto a caracteres delimitadores (punto, punto y coma, etc.). Una misma palabra podrá dar lugar a numerosos significados, pero esto no supone un gran inconveniente porque en pasos posteriores del análisis lexicométrico, las palabras o formas gráficas -como veremos más adelante- son recuperadas en su contexto de producción.

El alfabeto. Se definen por defecto, todos los caracteres del teclado como caracteres del alfabeto del lenguaje en el cual está escrito o transcrito el corpus. En el ejemplo utilizado en esta presentación, el corpus está compuesto por las respuestas de los 60 niños entrevistados y los delimitadores. La longitud del corpus está dada por el número de formas gráficas o palabras distintas. En nuestro caso es de 4070 palabras o formas gráficas.

La forma gráfica, el vocabulario y los delimitadores. Una cadena de caracteres acotada

por dos espacios o delimitadores como: punto, coma, dos puntos, puntos suspensivos, signos de interrogación, etc., se llama ocurrencia. El conjunto de las palabras o formas gráficas de un corpus constituye el vocabulario.

Particiones del corpus. Las producciones lingüísticas individuales pueden ser agrupadas en textos colectivos. Esto supone una partición del corpus en base a ciertos criterios, que se ponen a prueba mediante procedimientos lexicométricos que veremos más adelante. En el estudio con el que ejemplificamos la aplicación del método lexicométrico, las producciones individuales son respuestas dadas por los niños a cuatro preguntas de la entrevista. Las agrupamos en los siguientes textos, considerando en cada caso una variable distinta:

- según el nivel escolar, en: preescolar (GP), primer grado (G1), cuarto grado (G4);
- según el grupo social, en: marginado (A), medio (B) y
- según el género, en: masculino (M), femenino (F).

Decimos que la variable nivel escolar adopta tres modalidades, en tanto que grupo social y género adoptan dos modalidades respectivamente. En el primer caso, el corpus se particiona en tres textos colectivos, mientras que en los siguientes, en dos. Debido a nuestros objetivos en este estudio, establecimos como variable activa el grado escolar y como variables ilustrativas el grupo social y el género.

Variables activas e ilustrativas. Para evaluar las variaciones léxicas según variables de los individuos (o de la situación; por ejemplo, cortes temporales, condiciones de la producción lingüística, etc.), en los AFC es posible considerar variables activas y variables ilustrativas. Se consideran sólo dos variables activas a la vez:

- una siempre es la variable “palabras”;
- la otra puede ser, según los objetivos del estudio, o bien los individuos, o bien una variable de los individuos o de la situación de producción lingüística.

Las otras variables de los individuos o de la situación cuya relación con el léxico de los individuos se desee explorar pueden ser consideradas como variables ilustrativas. Las variables activas son las que definen la conformación de los planos factoriales sobre los que se visualizan los resultados. Las variables ilustrativas son proyectadas sobre los planos, sin incidir en la ubicación de las palabras y de las modalidades de la variable activa (en nuestro caso, las modalidades de la variable activa son PG, G1 y G4 para nivel escolar, M y F para género, etc.). Más adelante informaremos cuáles son los

indicadores a partir de los cuales se establece, a partir de los resultados de los procesamientos, si las modalidades de una variable activa y los de una ilustrativa tienen relevancia estadística.

Nota importante: En el caso de que una de las dos variables activas tenga sólo dos modalidades, el AFC se limitará a separar los individuos en dos grupos, según esas modalidades.

Glosario de las formas de un corpus (procedimiento MOTS, término francés para “palabras”). Las formas gráficas o palabras se pueden listar en orden alfabético o de frecuencia. Ambos listados pueden resultar útiles para fines distintos. Por ejemplo, la lista alfabética permite apreciar la presencia de términos en una misma familia de palabras, en tanto que el listado por frecuencias puede ser útil a la hora de establecer el umbral o filtro de frecuencia y de identificar las palabras gramaticales y las palabras llenas de mayor presencia en el corpus (aunque sin informar acerca de su distribución entre los individuos).

Umbral (procedimiento MOTS): establece un corte en el corpus que es necesario para el análisis. Como lo que interesa en un abordaje estadístico no es establecer la regularidad del vocabulario de un discurso individual sino la regularidad compartida por un grupo de personas, lo que es común a grupos, es preciso establecer un punto de corte entre lo que consideramos discurso compartido y simple variabilidad individual. El umbral se establece a fin de considerar para el Análisis Factorial de Correspondencias un subconjunto de las palabras distintas del corpus: aquellas que se registran al menos tantas veces como indica el umbral. En el estudio que tomamos como ejemplo, definimos un umbral de frecuencia = 10. Como no existen criterios para el establecimiento del umbral, se buscan los umbrales probando distintos valores (Crego, 1998). Esto es relativo a cada corpus. Así por ejemplo un umbral demasiado alto o demasiado bajo, en ciertos corpus, puede que no permita la delimitación de diferencias léxicas.

Análisis Factorial de Correspondencias (AFC) (procedimiento CORBIT, por “Correspondances Binaires Textuelles”). El programa SPADT utiliza el Análisis Factorial de Correspondencias, que se caracteriza por operar con dos variables. Dicho análisis permite estudiar la asociación entre modalidades de dos variables cualitativas definidas sobre una misma población de individuos. Opera a partir de una tabla de contingencia, que cruza las modalidades de las dos variables. En la tabla de contingencia las filas y columnas representan dos particiones de la misma población y ambas juegan en el problema un papel

análogo. En el caso de la lexicometría, una de esas variables es la variable palabras (todas las palabras distintas cuya frecuencia en el corpus es igual o superior al umbral de frecuencia). El término AFC surgió en 1962, de unos estudios sobre lingüística en los que J. P. Benzécri pretendió ofrecer un método inductivo eficaz para estudiar la distribución de una palabra. En dichos estudios definió la distribución Chi^2 como distancia distribucional. Remitimos al lector a Baccalá y Montoro (2008) para una explicación matemática de esta distancia. Siguiendo a las autoras, sólo señalamos aquí que la distancia Chi^2 es particularmente adecuada para apreciar proximidades entre perfiles filas y columnas, pues calcula la semejanza (en términos de distancia) entre dos filas, *ponderándola* según la importancia de las columnas. Como en una tabla de contingencia el papel de los perfiles fila o columnas es totalmente simétrico, ocurre algo similar para la distancia entre perfiles columna.

Nota importante: El Análisis Factorial de Correspondencias es una técnica de la estadística descriptiva no inferencial, cuyos resultados no admiten conclusiones que vayan más allá del colectivo observado.

Tabla Léxica. La tabla léxica contiene la frecuencia con la cual cada forma gráfica o palabra es empleada por cada uno de los individuos. Es una tabla de contingencia que tiene por filas los individuos y por columnas las palabras distintas en el corpus. Se trata de una tabla léxica parcial cuando sólo incluye las palabras retenidas después de la aplicación del umbral de frecuencia. En el estudio con el que ilustramos la aplicación, la tabla léxica (parcial) tiene 60 filas (porque son 60 los niños) y 71 columnas (porque son 71 las palabras distintas en el corpus total retenidas tras aplicar un umbral=10). El AFC aplicado a la tabla léxica (parcial) proporciona planos factoriales que permiten una visualización de las similitudes entre frecuencias de palabras entre sí, y de individuos entre sí. Entre filas y columnas se registran asociaciones baricéntricas. Es decir, a una palabra estarán asociados los individuos que la usan con mayor frecuencia. A partir de los agrupamientos de palabras observados en los planos resultantes del AFC de la tabla léxica, asociables al nivel escolar, sector social o género de los individuos, se pueden definir textos colectivos y continuar el análisis con una Tabla Léxica Agregada.

Tabla Léxica Agregada (procedimiento TALEX, por Tabla Léxica Agregada). Es una tabla de contingencia que tiene por columnas los textos reagrupados (por nivel escolar; sector social; género; o cualquier variable que se considere) y por filas las palabras o formas gráficas distintas. En el estudio con el que ilustramos la aplicación, la tabla léxica agregada

según la variable grado escolar tiene 3 columnas (correspondientes a preescolar, primer grado y tercer grado) y 71 filas (porque son 71 las palabras distintas en el corpus total retenidas tras aplicar el umbral). El AFC aplicado a la tabla léxica agregada proporciona en los planos factoriales una visualización de las similitudes entre perfiles de frecuencias de palabras y de asociaciones de éstos con las partes del corpus (textos reagrupados).

Concordancias (procedimiento CORDA). Este procedimiento permite el listado de los fragmentos del texto en los que se encuentran las palabras o formas gráficas importantes o características definidas estadísticamente u otras palabras que resulten de interés en el análisis. La forma gráfica o palabra se denomina “forma - polo” y el conjunto de fragmentos que la acompañan se denominan concordancias de la forma gráfica o palabra.

Formas o palabras características (procedimiento VOSPEC, por “Vocabulaire spécifique”). El procedimiento VOSPEC ofrece diversos resultados, que permiten describir la especificidad léxica de los textos reagrupados. Uno de ellos, denominado “formas o palabras características”, asocia a cada forma o palabra un valor test que mide la diferencia entre las frecuencias de la forma en la parte del corpus (por ejemplo, en el texto colectivo formado por las respuestas de los 20 niños de primer grado) y la frecuencia de la misma forma en la población general (en las respuestas de los 60 niños, cursen preescolar, primero cuarto grado). Dicho valor está “normalizado” por lo que se toman como características aquellas formas con un valor test (en valor absoluto) mayor a 1,96 (por lo que podremos descartar con una certeza (es decir, de probabilidad de equivocarse) del 5% que esa palabra esté allí por azar).

Respuestas modales o características (procedimiento VOSPEC, por “Vocabulaire spécifique”). Hasta este momento, hemos caracterizado un texto o sus partes por sus formas gráficas, o palabras. En estas operaciones hemos ignorado totalmente el contexto de una palabra o forma gráfica y el orden de las palabras o formas en las respuestas individuales. Otro de los resultados brindado por el procedimiento VOSPEC, la selección computacional de respuestas modales (completas), permite incorporar el contexto de producción de las palabras en el análisis lexicométrico, mediante el ordenamiento por su tipicidad de las respuestas más características de cada una de las modalidades de las variables de los individuos o de la situación. Este procedimiento adquiere especial importancia en las ciencias sociales ya que hace posible la interpretación del significado de las palabras en su contexto de producción. Se consideran en este caso las producciones lingüísticas originales de los individuos. En nuestro estudio, éstas son las transcripciones

digitalizadas de las respuestas originales pronunciadas por los sujetos entrevistados. El procedimiento VOSPEC ofrece dos criterios de selección de las respuestas características o típicas.: para respuestas largas se utiliza la distancia Chi^2 y para respuestas cortas, el criterio de valor test.

Distancia Chi^2 . La aplicación de este procedimiento (Lebart, 1982) permite extraer respuestas “reales” tales que su vocabulario sea representativo del vocabulario específico de cada modalidad de la variable en cuestión. Para ello, dado un grupo de individuos (por ejemplo: los niños de primer grado) se calcula el perfil medio de las respuestas del grupo a partir de los perfiles léxicos de las respuestas de los individuos que lo componen (es decir, de las respuestas completas dadas por todos los niños de ese grado escolar a las preguntas de la entrevista en las que focaliza el análisis) y se consideran como características de un grupo las respuestas más próximas a ese perfil medio. La proximidad entre respuestas y texto colectivo se mide en distancias Chi^2 . Dicha distancia es la utilizada en el Análisis Factorial de Correspondencias. Esta aplicación lista las respuestas completas de los individuos del grupo por su tipicidad en un orden decreciente.

Criterio del valor test. Se pueden también seleccionar las respuestas características siguiendo otro criterio, el criterio de valor test medio. Atribuye a cada palabra la media de los valores test de las palabras o formas que componen el texto. Las palabras con valor test medio más alto serán las características. Este criterio se utiliza sobre todo con respuestas cortas.

Plano Factorial: El AFC ofrece una representación conjunta de la nube de puntos fila y la de puntos columna. Para ello efectúa la proyección de las dos nubes sobre subespacios de dimensión reducida, pero manteniendo la máxima dispersión. Transcribimos la explicación que brinda de Torres Curth (2005: 59) al respecto: “El AFC proporciona una representación gráfica bidimensional de los elementos a describir denominada plano factorial (Crivisqui, 1993), en el cual se proyectan las modalidades de las variables definidas y los individuos. El plano está definido por dos ejes ortogonales (ejes factoriales), el primero representa la dirección de máxima dispersión de los datos. Las direcciones de los ejes factoriales están determinadas por las modalidades de las variables que más diferencian a los individuos, de modo que cada eje puede ser caracterizado por las modalidades que más contribuyen a su constitución, es decir que se encuentran por encima de la contribución media de todas las modalidades”.

6. Pasos para implementar la lexicometría con el programa SPAD.T

Una vez presentados algunos de los conceptos básicos, nos referiremos mediante la presentación de ejemplos basados en el estudio antes mencionado a:

- la carga de los ficheros,
- las instrucciones para activar los procesamientos y
- las salidas de los procesamientos.

Carga de los ficheros: cómo ingresar los datos

Fichero o archivo de respuestas textuales (lo llamaremos F1). Se lo prepara con un editor de textos (word, open office writer, etc.). Una vez que se lo completa, se selecciona el texto completo y se lo pasa enteramente a minúsculas, se lo guarda como txt o texto con tabulaciones, y se lo copia en una carpeta dedicada exclusivamente al procesamiento.

Este archivo contiene en forma secuenciada todas las respuestas o producciones de todos los individuos que conforman el corpus. Recomendamos volcar en este archivo el conjunto de la producción que se pretende analizar (por ejemplo, las respuestas a *todas* las preguntas de la entrevista, aunque en un primer paso sólo se analicen dos de ellas, que corresponden a un mismo objetivo o núcleo temático).

Se introduce cada individuo con ---- seguidos de dos espacios libres como máximo y luego el número de sujeto. La idea es que el programa “lee” el identificador del individuo considerando 8 lugares: 4 son los guiones y sólo lee 4 espacios más, de modo que si hubiese más de 99 sujetos, no podría dejarse más que un espacio en blanco.

La respuesta del individuo número 1 a la primera pregunta se introduce a continuación de un enter tras la identificación del individuo. En cambio, las respuestas a cada una de las siguientes preguntas se introducen luego de dos enter y 4 signos aditivos +++++

Si se cargan las respuestas de una entrevista, como en el ejemplo que aquí presentamos, no se incluyen las preguntas del entrevistador, sino sólo las respuestas de los entrevistados.

Cuando para un individuo no hay respuesta a una pregunta, debajo de los 4 signos aditivos o guiones correspondientes se escribe 9999

Al terminar, después de escribir la última respuesta del último sujeto, se introduce un

enter y luego se pondrán 4 signos igual ====

Es necesario homogeneizar la transcripción de los mismos términos. Para evitar la presencia de formas gráficas distintas para una misma palabra o expresión, es conveniente apelar a las convenciones ortográficas u otros acuerdos más puntuales al interior del equipo. Esto se debe a que el recuento se realiza sobre la ocurrencia de cada forma gráfica, de modo que “mi” y “mí” conforman formas distintas, tal como sucede por ejemplo con “eh”, “ehh” y “eeeh”. Esto lleva a que los investigadores deban decidir exactamente cómo desean transcribir las expresiones verbales de modo que las variaciones en las formas gráficas en el corpus sean intencionales y no producto de errores o descuidos.

Nota importante: cada individuo tiene que tener el mismo número de respuestas y éstas deben encontrarse en el mismo orden en todos los individuos. En caso de que esto no se cumpla, el procesamiento no podrá realizarse, informando error (“analyse stoppée”).

Ejemplo confección Fichero 1

Fragmentos extraídos del **fichero de respuestas textuales (F1)** utilizado en el estudio de las concepciones de aprendizaje de la escritura de niños de preescolar, primero y cuarto grado de la escuela básica.

El número total de preguntas analizadas fueron 22, de modo que para cada sujeto tenemos 22 respuestas, la primera introducida abajo de la identificación del individuo y las otras 21, abajo de una línea libre y +++. En el ejemplo sólo mostramos las respuestas de dos individuos a dos preguntas:

- ¿en tu casa alguien escribe?...
- ¿en qué piensa antes, durante, cuando borra y relea lo escrito?

Al interior de la respuesta de un individuo a una pregunta, el punto y aparte indica que entre una y otra emisión del niño hubo una intervención (pregunta, comentario, manifestación de acuerdo o de aliento) por parte de la entrevistadora.

```
---- 1
eh.
eh... mi hermana jimena.
para estudiar.
un día fui a su casa.
sí.
mi mamá.
eh... algunas veces.
y... para mandarle a mi papá.
```

por si sí ... por si viene o no viene y si se queda tiene que trabajar para que nosotros tengamos comida.

++++

acá está pensando...

y, haciendo cartas.

y, escribir letras.

eh.

eh... eh... haciendo... eh escribiendo un árbol.

para su papá.

eh... eh... corazones.

eh... pensar en la cabeza.

borrando.

porque le salió mal.

los corazones.

porque tiene ojos.

las letras.

eh la letra a.

las demás letras.

para aprender a escribir.

eh... que mi papá se dé cuenta.

que yo escribí letras.

su mamá.

cuando está borrando.

porque, porque la mamá le dice que borró y ella le dice que sí.

eh, una letra.

---- 2

acá, mi papá me enseñó a escribir tarea, todo eso, entendés?

y acá dice amor, esto lo escribí yo, esto lo escribió papá, esto

lo escribí yo y esto también.

todo esto lo escribí yo.

sí.

sí, papá me enseña, me enseña a hacer tarea, tarea cómica, tarea

clásica, y yo no sé hacer.

sí.

++++

9999

(continúa con las respuestas de los demás individuos)

====

Fichero de etiquetas (lo llamaremos F2). Se lo prepara con un editor de planilla de cálculo (excel u otros). Una vez que se lo completa, se eliminan todas las marcas de formato (colores, bordes, etc.) y se lo guarda como texto o txt con tabulaciones, y se lo copia en una carpeta dedicada exclusivamente al procesamiento.

En la primera columna se sitúa el identificador y cada columna siguiente da cuenta de una de las variables de los individuos (en nuestro caso: nivel escolar, sector social, género) o de la situación si ésta fuese objeto de análisis (por ejemplo: momento

temporal, modo de administración o de resolución de la tarea, etc.). En cada fila se completa un individuo, siguiendo exactamente el orden en el que aparecen en el fichero de texto (F1), y en cada columna se le asigna un número (con la función de etiqueta) que informa la modalidad que el individuo toma para esa variable.

Es útil preparar de antemano un sistema claro de denominación de modalidades de variables. En cuanto a los números para indicar las modalidades de cada variable es necesario utilizar una numeración correlativa comenzando a partir de 1. En el caso de no proceder así, el procesamiento se interrumpirá. En cuanto a las etiquetas con las que las modalidades aparecen en las salidas de los procesamientos, éstas pueden tener hasta 4 caracteres cualesquiera.

Ejemplo de confección del sistema de denominación de las modalidades de las variables

En el ejemplo que presentamos, el sistema de denominación de modalidades de variables fue:

Variable	Modalidad	Número utilizado en el fichero F2 para designar la modalidad	Etiqueta según la cual la modalidad aparece en los planos factoriales y demás salidas de los procesamientos
Sector social	Marginado	1	A
	Medio	2	B
Nivel escolar	Preescolar	1	GP
	Primer grado	2	G1
	Cuarto grado	3	G4
Género	Masculino	1	M
	Femenino	2	F

Sugerencia: En las etiquetas que aparecen en las salidas de los procesamientos, según el caso puede ser conveniente que unos caracteres indiquen la variable en cuestión y otros, la modalidad. Ejemplo: GP (modalidad preescolar de la variable nivel escolar); A (modalidad marginado de la variable sector social); etc.

Ejemplo de confección del Fichero 2 (F2)

En cuanto al Fichero 2, en la primera columna se introduce cada individuo y en las siguientes se le adjudica la etiqueta que corresponde a cada variable. La fila 1 corresponde al individuo 01 y la segunda

columna al sector social, la tercera al grado escolar y la cuarta al género. El orden de las etiquetas debe seguir el orden utilizado en la carga de los ficheros de palabras (para cada individuo) y etiquetas (para las variables). Si confrontamos:

- el F2
 - la tabla inmediatamente anterior donde sistematizamos la denominación de modalidades y
 - el fragmento del F1 en el ejemplo,
- podemos saber que los dos individuos provienen de sectores medios, cursan preescolar y son niñas.

```
01 2 1 2 (individuo de sectores medios, primer grado, femenino)
02 2 1 2
03 2 1 1
04 2 1 2
05 2 1 1
06 2 1 2
07 2 1 1
08 2 1 1
09 2 1 2
10 2 1 1
11 1 1 2
12 1 1 1
13 1 1 2
14 1 1 2
15 1 1 2
16 1 1 1
17 1 1 1
18 1 1 1
19 1 1 1
20 1 1 1
21 2 2 1
22 2 2 1
23 2 2 1
(continúa).
```

Nota importante: Antes de comenzar el procesamiento, **abrir una carpeta** específica donde se colocan **F1 y F2** y se guardan todos los ficheros que se van generando (NO guardarlos donde lo haría por defecto, que es en el propio programa).

Instrucciones para activar los procesamientos

Debido a que distintas versiones del programa tienen apariencia algo distinta, preferimos brindar las indicaciones en forma predominantemente verbal, prescindiendo intencionalmente de insertar imágenes de la apariencia de la pantalla en cada paso. Solo insertamos a título orientativo unos pocos íconos.

1. **BASE DE DATOS TEXTUALES.** En menu superior: **Base** → **Importer** → **Données textuelles** → se abre una ventana central que dice “**Importation de données textuelles**” → tica **Nouveau** → se abre una ventanita que dice “**Nom de l’importation**”, en la que debe tipearse cualquier cadena de caracteres (esto es sólo porque así permite seleccionar el archivo de la carpeta en la que se encuentra) → se abre una ventana con una alerta, se tica **OK** → seleccionar en carpeta adecuada el fichero con todas las respuestas textuales (F1) con doble click.

En **Importation du fichier des données textuelles**: ajustar el número de preguntas abiertas → **Nombre de questions ouvertes** (se puede identificar cada una de ellas con una etiqueta, aunque no es necesario) y de individuos → **N premiers individus**. En ambos casos, tiene que coincidir con el número de individuos y preguntas que componen el texto en F1 y los individuos en F2.

Exécuter → **Enregistrer** (=guardar) como base (terminación “.sba” en carpeta adecuada. Por ejemplo: texto.sba) → **Editer resultats**.

En **Archivage de donnés textuelles**, controlar que esté la totalidad de individuos y que el texto cierre con =====. → **OK** → **Fermer** (=cerrar).

2. **BASE DE ETIQUETAS.** En menu superior: **Base** → **Importer** → **Importer Fichier ASCII** → **Nouveau** (tipear cualquier caracter que no haya sido tipeado antes). Ahora se importa F2. El archivo tiene formato “txt con tabulación” (= “Tab delimited”). Seleccionar el archivo y marcar **Ouvrir**.

Mantener todos los comandos tal como están marcados. Verificar que en **Identificateur du texte** figure la opción **Aucun**.

En **La première ligne contient le libellé des variables**, poner **X** si la primer línea del F2 especifica el nombre de la variable. → **Suivant** (= siguiente)

Seleccionar en **type** la opción adecuada del menú descolgable: si la variable usa números como etiquetas, se elige “**nominale**”; cuando tiene números como frecuencias, se elige “**continue**”; cuando no se la quiere considerar, se elige “**abandon**”.

No es necesario completar las otras columnas de la tabla (libelle, cod.id, modalités).

Exécuter → **Enregistrer** como “**eti.sba**” → **Editer resultats** (para controlar que el número de individuos y variables sea el correcto) → **OK** → **Fermer**.

3. EDITAR LA BASE DE ETIQUETAS.

4. Menú superior: **Base** → **Editer base** (se selecciona la base que se creó a partir de F2: eti.sba). Para cada variable se cargan las modalidades y sus etiquetas. Para ello, posicionarse sucesivamente en cada columna a la derecha de la de “Libelle” y hacer lo siguiente:

- ticar en la barra superior horizontal de opciones en “**var**”. Se abrirá una tabla, en la que se verifica que el tipo de variable sea el adecuado así como el número mínimo y máximo que adopta la variable.

- posicionados en la primera variable de esa tabla, ticar “**mod**” en el menú superior. Se abre una nueva tabla y en la columna izquierda llamada “Libel” se coloca el nombre de las modalidades, de acuerdo a lo indicado en *Fichero de etiquetas* (ver p. 20). Luego de completar la primera variable, ticar en el **diskette** para guardar cambios. Ticar nuevamente en “**var**” y repetir el procedimiento para la fila siguiente (que indica la siguiente variable) → “**mod**”. Finalmente se tica en “**val**”, se graba y se cierra.

5. **BASE YUXTAPUESTA (TEXTO Y ETIQUETAS)**. Menú superior: **Outils** → **Juxtaposition**. De la carpeta correspondiente se selecciona “texto.sba” (creada a partir de F1). Posicionarse en el ícono

en la ventana central, y con doble click aparece una nueva ventana. En esa ventana se tica en “**ouvrir**” en la ventana superior derecha y se selecciona del directorio correspondiente eti.sba (creado a partir de F2).

Se tica abajo en **Base juxtaposée** y se la guarda en directorio correspondiente como “**YUXT.SBA**”.

Luego en **Paramètres** → **Mode de juxtaposition** → **Union**.

En **Indicateur de comparaison des individus** ticar en “**sur identificateur**”

Dictionnaire → **court**

Nombre d’individus dont on édite les données → **Tous**

Enregistrer en ventana central (cuadrante superior derecho) → **OK**.

Ticar en **Méthode** → **Exécuter méthode** → **oui** → se graba como FIL en directorio habitual: **YUXT.FIL** → **Enregistrer**. Aparece una ventana que pide un título adicional pero no es necesario poner nada. Ticar **OK**.

Ticar en ícono



en ventana central y fijarse en todas las salidas que los datos que aparezcan sean los correctos (número de variables e individuos).

Salir del Spad. Aparece una ventana “**Voulez-vous quitter SPAD?**” (= ¿Quiere salir del SPAD?) → **Oui**. Volver a entrar al programa.

En la ventana superior ticar **Base** → **Editer base**. Seleccionar la base **YUXT**. Aparece una ventana: “**Editeur de données YUXT**”.

6. **APLICACIÓN DE MÉTODOS**. Salimos del Spad. Aparece una ventana “**Voulez-vous quitter SPAD?**” (= ¿Quiere salir del SPAD?) → **Oui**.

Volvemos a entrar al programa. Sobre el ícono “**BASE**” ticar dos veces en el botón izquierdo del ratón (mouse) para poder seleccionar la base. Seleccionar la base yuxtapuesta (**YUXT.SBA**, no la **FIL**).

7. **INSERTAR MÉTODOS**. Colocar el cursor sobre la base, apretar botón derecho, ticar en **Insérer Méthode** que aparece en el descolgable y se crea un cuadrado gris vacío.

8. **MOTS**. Se selecciona de ventana a la izquierda: **Méthodes** (ticar en la flechita para poder seleccionar el análisis que se quiere hacer) → **Analyses Textuelles** (figura al final de la lista) → “**Mots**”.

Ticar sobre el cuadradito del método (en este caso, “**Mots**”) y arrastrarlo con el ratón hasta el cuadradito de color gris de la ventana central:



Ahora ticar dos veces en el método elegido (botón izquierdo), para seleccionar los parámetros del análisis.

Se seleccionan las variables textuales (Pantalla: **Sélection des Données Textuelles – Création du vocabulaire des mots**): se pasan al cuadro de abajo las preguntas que se quiera procesar ticando sobre la flecha dirigida hacia abajo.

(Para abrir las siguientes pantallas, pulsar en las pestañas de abajo).

En la ventana **Individus** se especifican los individuos: si se trabaja con todos, se

selecciona “**Tous**”, como está por defecto. Si se quiere trabajar sólo con una parte de ellos, explorar las otras alternativas. Dejar la otra información como está.

En la ventana **Vocabulaire des mots** se especifica el vocabulario de la siguiente manera:

- en “**Gestion de la casse du texte**” → sans changement (=sin cambios).
- en “**Les séparateurs**”: en el caso de textos orales, pasar todos los separadores “**De phrases**” excepto el punto a separadores de palabras o “**De mots**”. Hacerlo con la función de cortar y pegar. En el caso de que el texto presente más clases de separadores que los que muestra la pantalla (por ejemplo ¿ ó ¡), incluirlo donde corresponda.

Ticar **OK** → **Éxecuter Méthode**.

9. **UMBRAL O FILTRO, EN MOTS**. En la salida de resultados, ticar en ícono con letras ABC para establecer el umbral de frecuencia. Para ello:

En menú superior ticar **Edition** → **Sélection par filtre** → **Fréquence** → menor o igual (< o =) → en casillero superior a la derecha tipear el número elegido como umbral; de este modo deja las palabras con frecuencia = al número de umbral elegido) → **Ajouter** → **OK**.

En la ventana central aparecen automáticamente marcadas todas las palabras con frecuencia menor al umbral. En la barra horizontal superior de herramientas marcar la **X** de color negro, así se las borra (aunque se registran entre paréntesis). También se pueden marcar una a una con el cursor y luego suprimir con el botón derecho.

Ticar en menú superior: **Edition** → **Statistiques** y allí consultar y anotar los valores para **Vocabulaire Initial** y **Modifié**. Anotar en otro archivo los valores de nombre de mots y frecuencia total (son la cantidad de palabras distintas y totales para el vocabulario antes y después de la aplicación del umbral).

En casos muy particulares puede aplicarse un filtro de longitud de palabras antes del de frecuencia, para suprimir por ejemplo las palabras de una y dos letras de extensión.

Este procedimiento también permite corregir errores de tipeo. En la columna “**Vocabulaire modifié**”, posicionarse arriba de la palabra a modificar, con botón derecho del ratón, apretar “**Corriger**” y hacerlo sobre la palabra, luego “**Valider**”. Quedan marcados los cambios en rojo y las supresiones en azul. Es conveniente realizar estas correcciones antes de aplicar umbrales de frecuencia y/o longitud.

Ticar en “**Valider**” → Dar al ícono del **diskette** para guardar → Guardar como página excel lo que hay en el ícono que lleva el símbolo “**excel**” (el que está más a la derecha), donde lista el vocabulario inicial por orden de frecuencia y alfabético.

Nota importante: a partir de aquí, guardar todas las salidas EXCEL pues son de gran utilidad para los informes.

Para insertar un método sucesivo, posicionarse en el último método ejecutado y ticarlo una vez con el botón derecho. Luego → **Méthode** → **Insérer méthode**.

10. **VOSPEC.** Después de “**Mots**” insertar “**Vospec**” (vocabulario específico).

Ticar dos veces en el ícono “**Vospec**” con el botón izquierdo del ratón para seleccionar las “**Variables**” y los “**Paramètres**”.

“**Variables**”: bajar con la flecha aquella que queramos considerar como “variable activa”; las demás se quedan donde están.

“**Paramètres**”: Dejar todo por defecto. **OUI** → Dar **OK**.

Luego ticamos en “**Vospec**” con el botón derecho → “**Éxecuter méthode**”. Para ver los resultados, ticar en el ícono del Excel. Guardar esta salida de datos como documento de excel: “**VOSPEC-Modalidad. Excel**”.

Si interesa conocer el vocabulario específico según cada variable, se repite el procedimiento para cada una de ellas. En nuestro estudio, lo hacemos tres veces: para grado escolar, de modo de conocer el vocabulario específico de los niños de preescolar, primer grado y cuarto grado; para género, para conocer el vocabulario específico de niñas y de varones; y para grupo social, para conocer el vocabulario específico de los niños que provienen de sectores socioculturales medios y el de niños de sectores marginados.

Nota importante: Como sucede con todos los resultados de la lexicometría, el vocabulario en cuestión es específico de los grupos estudiados, y no de todo el universo de referencia.

11. **TALEX.** Después del VOSPEC, aplicar el TALEX (el procedimiento para empezar es el mismo que antes: ticar en el método anterior, indicar insertar nuevo método, elegirlo y arrastrarlo con el ratón, etc.).

Ticar en **Paramètres**. Se abre una ventana que dice “**Tableau léxical de contingente**” En “**Tableau léxical de contingence**” se abre automáticamente la selección de variables: seleccionar todas las variables y mandarlas a la casilla de abajo. Luego ticar en la pestaña donde dice “**Paramètres**” y marcar “**OUI**” en todo → **OK**.

Luego ticar en el cuadradito donde dice “**TALEX**” → “**Éxecuter méthode**”.

Guardar los resultados de este análisis en la carpeta que corresponda.

En la ventana **MOTS.FIL** ticar **Filière** → **Sauvegarder filière**.

12. **CORBIT**. Después del TALEX aplicar el “**CORBIT**” (es el Análisis Factorial de Correspondencias Simples o Binarias Textuales). El proceso para introducir el nuevo análisis es el mismo de siempre (ver más arriba).

Ticar dos veces en el método → se abre ventana **Analyse de Correspondances Binaires**. En la ventana izquierda seleccionar qué variable se va a tomar como activa y cuál /es como ilustrativa/s. Para ello, ticar en “**TYPE**” y marcar lo que se quiere. Dejar la ventana derecha (mots) como está.

En “**Paramètres**” → “**Fichier pour Application tableur**”. Pulsar en “**Options**” → “**Options pour application tableur**” → “**Résultats pour les lignes**” → “**OUI**” → **OK**.

Dar a **OK** en la pantalla general.

De nuevo, dar en “**Exécuter Méthode**” (como siempre).

Aclaración: en el CORBIT, el programa denomina “individuos” a las palabras y “frecuencias” a las modalidades de la/s otra/s variable/s (como grado escolar). Aclaremos esto porque en los demás procedimientos los individuos coinciden por lo general con las personas.

Se obtienen 4 archivos con las salidas, en 4 recuadros.

- Ticar en el cuarto recuadro, el excel, que brinda todos los resultados del AFC, y guardarlo en la carpeta adecuada. Atender especialmente: Contribuciones de frecuencias activas en hoja “**corbi-2**” (o modalidades de la variable activa, en el ejemplo, grado escolar) y contribuciones de individuos activos en hoja “**corbi-4**” (modalidades de la variable palabra). Se tienen en cuenta para la conformación de los grupos léxicos aquellas modalidades cuya contribución supera la contribución media. Ésta se calcula para cada variable dividiendo 100 por la cantidad de modalidades. En el ejemplo, para las frecuencias activas es 33% (100/3 grados escolares) y para los individuos activos (o

palabras) es 1,40 (100/71, pues son 71 las palabras distintas retenidas después de la aplicación del umbral). Para verificar la calidad de representación de las frecuencias activas contributivas y de los individuos activos contributivos en el plano factorial, hay que mirar los cosenos indicados en la hoja corbi-2 y corbi-4 respectivamente. En la lexicometría es habitual obtener una buena calidad de representación, de modo que la suma de los cosenos de los ejes que conforman el plano factorial para la modalidad contributiva en cuestión sea igual a, o supere 0,40 (señalemos que este valor es sólo una indicación que deberá ajustar el investigador en cada caso).

Ticar en el primer recuadro:



Ver histograma de los valores propios de los ejes, que también aparece en “**corbi-1**” del archivo excel.



Ticar en el segundo recuadro:

En la ventana que se abre entonces, ticar en “**Graphique**” → “**Nouveau**” → “**Sélectionner**” → en la ventana “**Préselections pour un nouveau graphique**”, marcar “**individus actifs**” (para que en el plano se proyecten todos los individuos); “**fréquences actives**” (para que se proyecten las modalidades de las variables activas) y “**fréquences illustratives**” (para que se proyecten las modalidades de las variables ilustrativas) → **OK**.

Se abre una ventana que muestra el primer plano factorial.

Ticar en el noveno recuadro superior (un cuadrado con puntos, “**sélection par cadrage**”) y con el ratón marcar toda la superficie del plano factorial. En el recuadro central, ticar:

Sélection par cadrage → **Intérieur du cadre** → **OK**.

Ticar en el recuadro superior **abc** y se verán las modalidades de las variables (palabras y etiquetas de todas las variables utilizadas, tanto activas como ilustrativas, en el ejemplo: PG, P1, P4; F, M; A, B).

Ticar en el sexto recuadro para visualizar otros planos factoriales. Según el histograma de valores propios de los ejes factoriales y de las contribuciones de las modalidades de las variables (estos resultados se consultan en el archivo excel de resultados del

CORBIT), puede ser conveniente observar los planos factoriales que resultan de los ejes 1 y 3, 2 y 3, etc.

Para grabar, ticar **Graphique** en el menú superior → **Enregistrer sous sauvegarde interne** y luego repetir la acción para grabar con formato **BMP: Graphique → Enregistrer sous → Fichier image BMP**. Ubicar en carpeta adecuada.

Se sugiere explorar los diversos comandos para mejorar la representación en el plano y extraer y comprender la gran cantidad de información disponible. Unas mínimas pistas: situando el cursor sobre las modalidades proyectadas en el plano y apretándolo de diversos modos, es posible desplazar la etiqueta que puede estar ocultando otras, ver los resultados estadísticos para esa modalidad y editar la etiqueta.

12. **TEXNU**. Una vez realizado el CORBIT, insertar método TEXNU. El proceso para introducir el nuevo análisis es el mismo de siempre (ver más arriba). Ticar dos veces en el método.

En “**Création d’une base mots/segments/variables**”, cuando en el desplegable aparece “**mots/segments**” ticar en la doble flecha para mandar todas las palabras a la casilla de abajo. Luego en el desplegable elegir “**variables numériques**” (son todas las variables que no son las palabras, en el ejemplo: grado escolar, sector social y género) y tocar en doble flecha para mandarlas abajo, a “casille inférieur” (éstas son las que en el siguiente paso consideraremos como ilustrativas).

Ticar abajo a la izquierda (**Nouvelle base**) y grabar la base en el directorio correspondiente como **TEXNU → OK**. Ejecutar método como siempre.

En la ventana supérieure, ticar **Filière → Sauvegarder filière**.

Cerramos el SPAD porque ahora cambiaremos de base.

Volvemos a abrir el programa.

13. **CORBI**. En pantalla central: **Filière → selectionner base →** selecciono **TEXNU.sba → Insérer méthode**. Lo seleccionamos desde **Análisis Factoriales** (NO de Análisis Textuales) y entre ellos, elegimos correspondencias binarias (**CORBI**). Realizar este paso es fundamental para poner a prueba la hipótesis nula, que consiste en la no diferenciación léxica de acuerdo a las variables consideradas. Este procedimiento opera sobre la tabla léxica de individuos por palabras. Realiza un análisis factorial de correspondencias simples sobre dicha tabla. Por lo tanto, las variables activas serán

individuos y palabras (en el ejemplo, los 60 niños y las 71 palabras). Se consideran todas las demás como variables ilustrativas (en el ejemplo, grado escolar, sector social y género).

En **Paramètres: Colonne**. En “**sélection de colonnes**”, se eligen las **Fréquences actives** (son las palabras): se bajan todas por medio de la doble flecha. Luego en “**sélection de colonnes**” se elige “**Nominales illustratives**” y se bajan todas del mismo modo.

En la ventana **Paramètres**, a la derecha: en **Fichier pour application tableur**, en **Options** marcar **OUI** (caso contrario, no se generará el archivo excel) →**OK** → **Exécuter**. Solicita guardar, hacerlo en carpeta correspondiente.

Luego se procede a mirar el fichero excel de resultados y generar los gráficos, solicitando que grafique frecuencias activas, individuos activos y variables ilustrativas, de modo similar a lo indicado en el paso 11 para CORBIT.

En el fichero excel, mirar en **Valeurs-Test des modalités illustratives en corbi-8**. Se considera que las modalidades cuyo valor test es igual o superior a 1.96 en valor absoluto presentan un léxico diferenciado de las demás, con una probabilidad de .05. Este archivo también informa en otras hojas las contribuciones de las modalidades de la variable individuos (que en este caso por lo general son las personas, en nuestro ejemplo, los 60 niños) y de la variable palabras. Como en el CORBIT, se considera aquellas con una contribución superior a la media.

Es decisión del investigador si avanza hacia el AFC aplicado a la tabla léxica agregada (como mostramos en el paso 11, CORBIT), o da cuenta de sus resultados en función del CORBI aplicado a la tabla léxica (paso 13). En uno u otro caso, se describen los grupos léxicos utilizando el VOSPEC.

14. **CORDA**. Este procedimiento permite identificar los fragmentos de todas las respuestas en las que están situadas palabras cuyas variaciones de uso interesan al investigador. De ese modo, el procedimiento hace posible visualizar cada palabra en su contexto original de aparición.

Se selecciona de ventana a la izquierda: **Méthodes** (ticar en la flechita para poder seleccionar el análisis que se quiere hacer) → **Analyses Textuelles** → “**Corda**”. Arrastrar el cuadradito del método con el ratón hasta el cuadradito de color gris de la

ventana central.

En la pantalla **Édition des Contextes des Mots**, en el cuadro superior se encuentran disponibles todas las palabras distintas del corpus analizado. Con las flechas, se pasan al cuadro de abajo las palabras que interesa investigar.

En la ventana **Paramètres en Edition des contextes avec identificateurs des individus** se tica “Oui”.

Ticar **OK →Éxecuter Méthode**. Abrir el fichero excel. Cada hoja del fichero brinda la información relativa a cada palabra seleccionada, informado todas las ocurrencias de la misma en el corpus, en su contexto de producción. En la última columna a la derecha figura el número de individuo.

7. Las salidas de los procesamientos en el estudio utilizado como ejemplo

Se presentan a continuación las salidas de los procedimientos estadísticos principales que sirvieron de base para el establecimiento de las diferencias léxicas y la elaboración de la descripción de dichas diferencias según la variable grado escolar. Los resultados de este estudio se encuentran en Scheuer, de la Cruz, Pozo, Huarte, y Sola, 2006, y pueden ser consultados como guía de interpretación de resultados y modelo de presentación de los mismos. En esta presentación diferenciamos tres etapas principales:

- En la primera etapa se pone a prueba la hipótesis de diferenciación léxica (procedimiento CORBI con base Texnu).
- En la segunda etapa y según los objetivos del estudio y la factibilidad de los datos, se profundiza el análisis de la diferenciación léxica (procedimiento CORBIT en SpadT).
- En la tercera etapa se obtienen las respuestas típicas (procedimiento Respuestas Modales). Otro procedimiento informa en fragmentos de respuestas, a partir de una palabra (procedimiento CORDA). A partir de estas respuestas se elabora la descripción de los grupos y luego su interpretación estableciendo conexiones con la teoría o enfoques de referencia.

Primera etapa: poner a prueba de la hipótesis de diferenciación léxica

Las salidas, después de aplicar el procedimiento MOTS, muestran la totalidad de

las palabras ordenadas alfabéticamente y según las frecuencias acumuladas. Informa también la cantidad total de palabras del corpus que se analiza y la de palabras distintas. Permite la aplicación del Umbral (ver Apéndice).

ORDEN ALFABÉTICO.

SELECTION DES MOTS

UMBRAL DE FRECUENCIA = 10
 TOTAL DE PALABRAS RETENIDAS = 4070
 TOTAL DE PALABRAS DISTINTAS = 586

Liste des mots par ordre alphabétique

Numéro	Mots employés	Fréquences	Longueurs
1	a	113	1
2	abajo	3	5
3	abecedario	1	10
4	abuela	8	6
5	abuelita	2	8
6	abuelo	5	6
7	acento	1	6
8	acentos	1	7
9	acordándose	1	11
10	acordó	1	6
11	acuerda	3	7
12	acá	6	3
13	adentro	1	7
14	adónde	1	6
15	ahora	7	5
16	ahí	20	3
17	aires	5	5
18	ajá	2	3
19	al	8	2
20	algo	26	4
21	alguien	3	7
22	alguna	16	6
23	algunas	2	7
24	alguno	1	6

Tabla 1. Lista de palabras por orden alfabético, donde se informa su frecuencia y la extensión de cada palabra, es decir, el número de caracteres que la constituyen. Por razones de espacio la tabla aparece incompleta, ya que el número total de palabras distintas es 71.

Liste des mots par ordre de fréquence

Numéro	Mots employés	Fréquences	Longueurs
448	que	224	3
575	y	162	1
290	la	125	2
296	le	125	2
433	porque	124	6
1	a	113	1
217	está	94	4
169	en	93	2
372	no	82	2
318	lo	79	2
193	escribir	76	8
323	mal	75	3
491	se	70	2
518	sí	62	2
555	una	53	3
412	pensando	50	8
398	para	48	4
499	si	48	2
487	salió	47	5
162	el	46	2
514	su	45	2
554	un	45	2
381	o	45	1
557	va	44	2
125	de	43	2
457	qué	40	3
71	carta	38	5
54	bien	35	4
295	las	35	3
441	puede	33	5
536	tiene	32	5
90	con	31	3
302	letras	31	6
325	mamá	30	4
301	letra	29	5

Tabla 2. Lista de palabras ordenadas por frecuencia. Por razones de espacio la tabla aparece incompleta, ya que el número total de palabras distintas es 71.

Mediante el procedimiento TALEX se obtiene la siguiente tabla de frecuencias brutas (filas= palabras; columnas= modalidades de las variables). Las celdas informan la frecuencia de la palabra en la fila en la partición del corpus indicado en la columna. Así, como muestra la siguiente tabla, en el conjunto de respuestas dadas por los niños de preescolar (GP), la palabra *carta* se registró 19 veces, en tanto que apareció 4 veces en las respuestas de primer grado (G1) y 15 en las de cuarto grado (G4).

Tableau de contingence

	GP	G1	G4	SMA	SME	M	F
a	37	37	39	43	70	46	67
ahí	5	9	6	14	6	8	12
algo	3	3	20	12	14	13	13
alguna	5	0	11	7	9	3	13
así	3	9	6	10	8	3	15
bien	12	7	16	18	17	26	9
carta	19	4	15	25	13	19	19
como	1	4	8	2	11	3	10
con	8	9	14	8	23	10	21
cosas	2	2	11	9	6	5	10
cuando	3	6	4	7	6	6	7
cuenta	2	3	7	5	7	3	9
cuento	3	13	8	12	12	13	11
cómo	3	5	17	13	12	14	11
de	6	21	16	20	23	15	28
después	7	6	4	8	9	8	9
eh	18	2	2	5	17	7	15
el	9	24	13	13	33	19	27
ella	4	12	3	10	9	1	18
en	28	34	31	48	45	47	46
es	3	4	4	4	7	8	3
escribe	2	13	10	14	11	10	15
escribiendo	4	4	4	4	8	3	9
escribir	14	40	22	42	34	32	44
escribió	5	6	13	8	16	10	14
estaba	2	6	5	7	6	4	9
estar	2	3	11	7	9	10	6
está	35	31	28	49	45	46	48
hacer	13	8	1	13	9	14	8
hizo	4	7	3	9	5	6	8
la	33	52	40	56	69	61	64
las	16	9	10	18	17	16	19
le	47	28	50	68	57	67	58
letra	6	9	14	17	12	15	14
letras	18	10	3	15	16	10	21
leyendo	5	2	7	7	7	6	8
lo	18	31	30	39	40	47	32
los	7	8	5	10	10	9	11
mal	22	20	33	46	29	43	32
maná	8	16	6	19	11	10	20
mi	2	8	4	10	4	3	11
no	22	35	25	46	36	26	56
nombre	4	10	1	7	8	5	10
o	3	10	32	13	32	17	28
otra	6	2	3	5	6	8	3
palabra	0	1	14	4	11	7	8
para	22	15	11	25	23	27	21
pensando	18	14	18	24	26	28	22
pensar	1	7	4	2	10	1	11
piensa	5	2	4	7	4	6	5
poner	1	1	13	11	4	7	8
por	2	3	9	5	9	4	10
porque	36	49	39	69	55	61	63
puede	6	8	19	19	14	14	19
que	54	82	88	126	98	95	129
quiere	2	4	5	5	6	6	5
qué	6	10	24	26	14	18	22
salió	19	14	14	21	26	28	19
se	18	18	34	29	41	28	42
si	2	8	38	19	29	25	23
su	17	19	9	24	21	15	30
sé	7	12	2	10	11	9	12
sí	26	17	19	31	31	31	31
te	9	5	3	3	14	6	11
tiene	7	11	14	12	20	12	20
todo	5	7	5	11	6	6	11
un	21	17	7	22	23	24	21
una	19	11	23	21	32	25	28
va	7	13	24	15	29	25	19
y	46	53	63	66	96	73	89
yo	2	8	2	6	6	3	9

Tabla 3. Tabla de contingencia que informa la frecuencia de las palabras según las variables consideradas en el estudio.

La Tabla 4 muestra que en los ejes factoriales 1 y 2, tienen valor test superior a 1,96, todas las modalidades de la variable Grado escolar y la modalidad M (masculino) de la variable Género.

Valeurs-Tests des modalités illustratives

Libellé	Effectif	Poids absolu	Distance à l'origine	Axe 1	Axe 2	Axe 3	Axe 4	Axe 5
grado								
GP	20	837,00	4,40085	-3,90	0,71	-2,69	0,13	0,11
G1	20	991,00	3,85357	-1,76	-2,24	0,54	1,19	-0,93
G4	20	1120,00	2,24998	4,47	1,44	1,53	-1,15	0,74
sector								
SMA	30	1435,00	2,04412	-0,71	0,29	-1,88	-0,49	-0,74
SME	30	1513,00	1,95783	0,67	-0,28	1,78	0,46	0,70
género								
M	30	1339,00	1,97788	0,18	2,11	-2,11	-0,58	-0,64
F	30	1609,00	2,02407	-0,15	-1,76	1,76	0,48	0,53

Tabla 4. Valores-test de las modalidades ilustrativas. Figura el número total de individuos por modalidad y los valores-test en cada eje (axe) de cada una de las modalidades de las variables ilustrativas.

La Figura 2 muestra que las modalidades G1 y G4 de la variable grado escolar son opuestas e interdependientes (lo que estaría indicando que aquello de lo que más hablan los niños de G1 es de lo que menos hablan los de G4 y viceversa). En cambio los niños de GP tienen un léxico independiente, hablan de otras cosas.

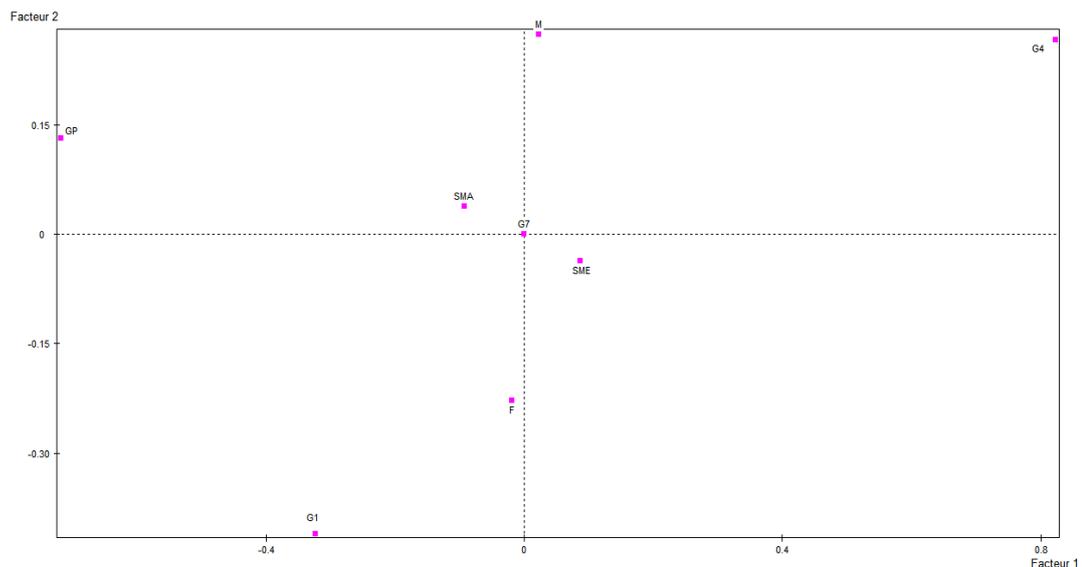


Figura 2. Plano Factorial del AFC sobre la tabla léxica. Proyección de las modalidades de las variables ilustrativas: grado escolar, género y grupo social.

La Figura 3 muestra además a los individuos (1 a 20= PG; 21 a 40= G1; 81 a 100= G4).

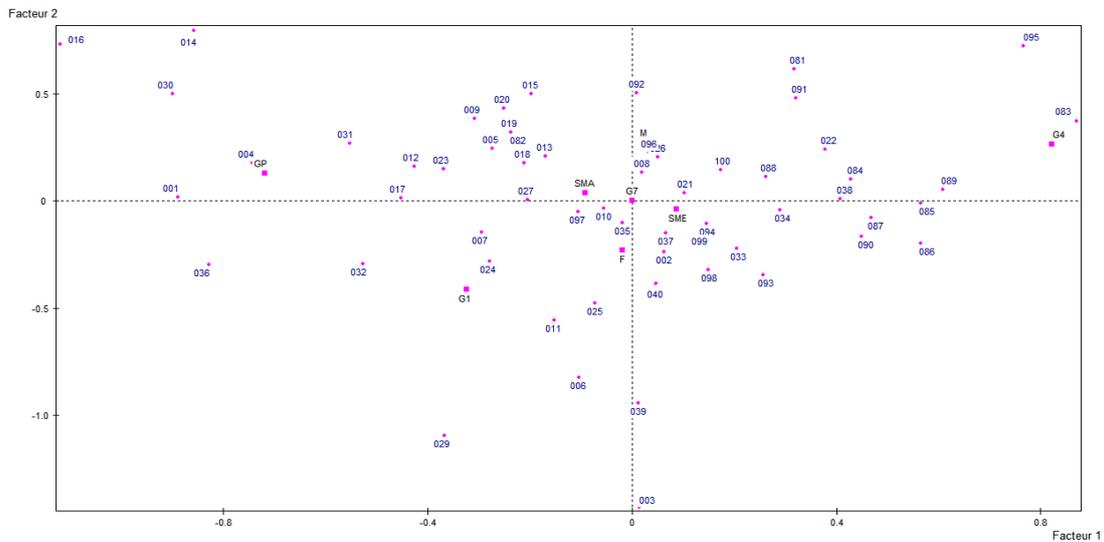


Figura 3. Plano Factorial del AFC sobre la tabla léxica. Proyección de las modalidades de las variables ilustrativas (grado escolar, género y grupo social) y de los individuos.

Segunda etapa: identificar grupos léxicos

Una vez comprobada la diferenciación léxica según las variables estudiadas (mediante el puntaje Valor test de las modalidades de las variables obtenido en la primera etapa del análisis), se avanza hacia un segundo análisis factorial de correspondencias simples. En este caso algunas modalidades de variables, según los objetivos del estudio y los valores test obtenidos, se consideran activas y por tanto junto con las palabras, contribuyen a la conformación de los ejes factoriales. Las otras variables operan como ilustrativas. Se considera que las modalidades activas tienen importancia estadística cuando su valor es igual o superior a la media. La contribución superior a la media de las modalidades de la variable grado escolar en nuestro ejemplo es 33,3 (resulta de $100/3=33,33$). La contribución superior a la media de las palabras de la variable palabra es 1.40 (resulta de $100/71=1,40$). Las Tablas 5 y 6 muestran las contribuciones de las modalidades de la variable grado escolar y de la variable palabras respectivamente.

Contributions des fréquences actives

Libellé de la variable	Poids relatif	Distance à l'origine	Axe 1	Axe 2
CP	28,39	0,18018	30,24	41,37
G1	33,62	0,11757	10,01	56,38
G4	37,99	0,15532	59,75	2,26

Tabla 5. Contribuciones de las modalidades de grado escolar (frecuencias activas) en los dos ejes considerados

Contributions des individus actifs

Identificateur	Poids relatif	Distance à l'origine	Axe 1	Axe 2
a	3,83	0,01008	0,28	0,22
ahí	0,68	0,05942	0,11	0,55
algo	0,88	0,64398	5,58	0,52
alguna	0,54	0,58805	1,63	3,04
así	0,61	0,13399	0,00	1,54
bien	1,19	0,08308	0,15	1,59
carta	1,29	0,32362	0,08	7,73
como	0,44	0,29926	1,25	0,20
con	1,05	0,02213	0,22	0,04
cosas	0,51	0,53101	2,65	0,27
cundo	0,44	0,07044	0,05	0,50
cuenta	0,41	0,17942	0,75	0,00
cuento	0,81	0,22030	0,00	3,39
cómo	0,85	0,38681	3,37	0,04
de	1,46	0,14251	0,05	3,84
después	0,58	0,11346	0,62	0,10
eh	0,75	1,40411	4,69	11,22
el	1,56	0,15482	0,28	4,05
ella	0,64	0,40834	0,88	3,36
en	3,15	0,00933	0,28	0,05
es	0,37	0,00338	0,00	0,02
escribe	0,85	0,24806	0,16	3,68
escribiendo	0,41	0,01434	0,05	0,02
escribir	2,58	0,16412	0,35	7,36
escribió	0,81	0,11107	0,90	0,05
estaba	0,44	0,10641	0,02	0,84
estar	0,54	0,40371	2,23	0,06
está	3,19	0,04537	1,20	0,54
hacer	0,75	0,62862	4,35	0,91
hizo	0,47	0,15208	0,44	0,56
la	4,24	0,02981	0,42	1,62
las	1,19	0,14762	0,79	1,87
le	4,24	0,06835	0,00	5,47
letra	0,98	0,05072	0,50	0,02
letras	1,05	0,52167	4,65	1,87
leyendo	0,47	0,16799	0,16	1,21
lo	2,68	0,02049	0,02	1,00
los	0,68	0,07193	0,50	0,02
mal	2,54	0,02418	0,28	0,65
mamá	1,02	0,20190	1,05	1,97
mi	0,47	0,25810	0,05	2,22
no	2,78	0,04014	0,47	1,26
nombre	0,51	0,58428	1,62	2,65
o	1,53	0,49358	7,77	0,03
otra	0,37	0,34202	0,40	1,69
palabra	0,51	1,30611	6,76	0,19
para	1,63	0,16862	2,13	1,29
pensando	1,70	0,03081	0,09	0,82
pensar	0,41	0,32917	0,00	2,53
piensa	0,37	0,17410	0,05	1,14
poner	0,51	1,00591	5,07	0,41
por	0,47	0,29625	1,44	0,02
porque	4,21	0,02176	0,63	0,57
puede	1,12	0,16381	1,87	0,04
que	7,60	0,00957	0,15	1,09
quiere	0,37	0,05362	0,13	0,14
qué	1,36	0,21274	2,98	0,00
salió	1,59	0,07308	0,70	0,92
se	2,37	0,05056	1,02	0,41
si	1,63	0,73841	12,41	0,02
su	1,53	0,13826	2,15	0,05
sé	0,71	0,38657	2,12	1,33
sí	2,10	0,09023	0,85	2,03
te	0,58	0,32647	1,41	0,99
tiene	1,09	0,02386	0,21	0,11
todo	0,58	0,03675	0,15	0,12
un	1,53	0,25528	3,83	0,36
una	1,80	0,07648	0,07	2,47
va	1,49	0,13194	1,98	0,09
y	5,50	0,00045	0,01	0,02
yo	0,41	0,49307	0,45	2,97

Tabla 6. Contribuciones de las palabras en los ejes considerados.

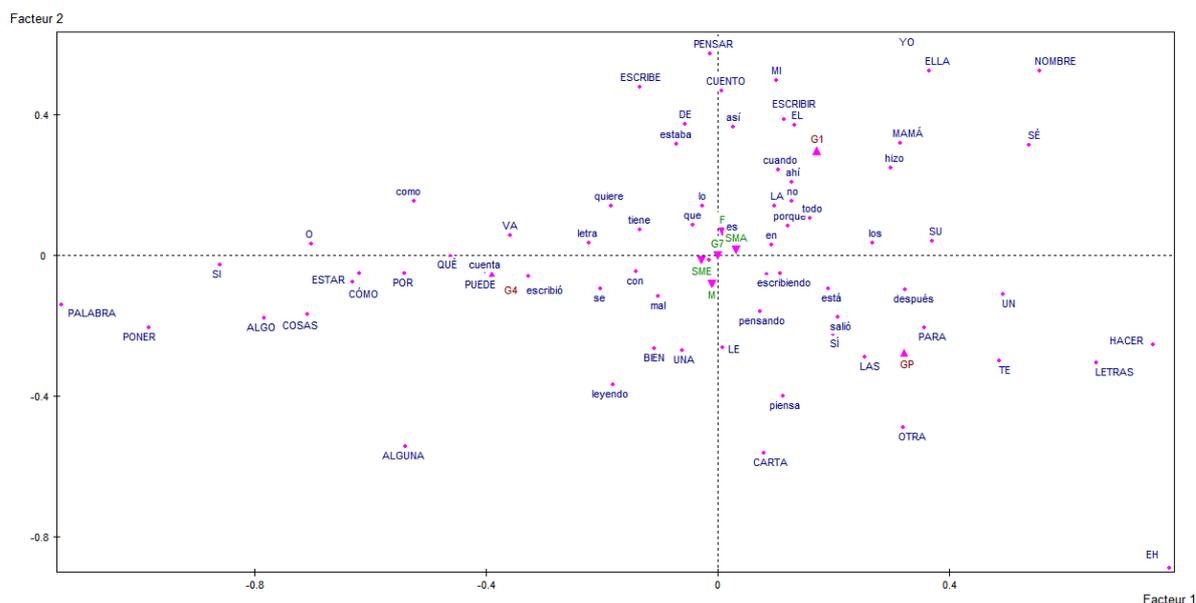


Figura 4. Plano factorial resultante del análisis factorial de correspondencias simples aplicado a la tabla léxica agregada. Las palabras con letra mayúscula y las siglas de las variables (GP, G1 y G4) indicadas en el plano, contribuyen por encima de la media a uno o los dos ejes factoriales.

Este segundo análisis factorial de correspondencias permite diferenciar los siguientes grupos:

Grupo 1: **carta, letras, otra, un, su, hacer, para, eh, sé, le, te, una.** La modalidad **preescolar (GP)** se asocia a este grupo.

Grupo 2: **nombre, mamá, ella, yo, mi, el, la, pensar, escribir, escribe, de, cuento, sé** y la modalidad **primer grado (G1)**

Grupo 3: **palabra, cosas, algo, alguna, poner, puede, va, estará, si, cómo, qué, o** y la modalidad **cuarto grado (G4).**

Tercera etapa: describir respuestas modales

Los resultados de los procesamientos de la primera y segunda etapa han permitido establecer que existen diferencias léxicas según las modalidades de las variables consideradas. En esta tercera etapa el análisis avanza con la aplicación del procedimiento de respuestas modales. Éste permite interpretar el significado del léxico o palabras que alcanzaron importancia estadística en su contexto de producción, es decir en las respuestas "reales" y completas de los sujetos entrevistados. Para ello, el

procedimiento estadístico selecciona las respuestas típicas o características correspondientes a cada modalidad. Dicho procedimiento ordena jerárquicamente las respuestas según criterio Chi². Para las modalidades de la variable grado escolar en la tarea seleccionada, tenemos las siguientes respuestas (en general se consideran en la descripción más de una respuesta porque se supone que la primera no es suficiente para describir el léxico que caracteriza a los grupos diferenciados), que se muestran en las Tablas 7, 8 y 9 (para preescolar, primer grado y cuarto grado respectivamente).

A continuación ilustramos, tomando las tres primeras respuestas típicas de cada modalidad de la variable grado. La columna de la izquierda indica la distancia Chi² de la respuesta, la columna de la derecha de la respuesta, indica el número que permite identificar al individuo que la produce.

Groupe d'individus : GP

Critère de classement	Numéro	Libellé de la réponse	Identificateur
0,561	1	una carta para su seño. estar pensando que le haga un dibujo. en una casa. porque le habrá salido mal la manzana. porque era redondita. porque eran redonditas y las hizo con puntitos. le dice a su mamá . mamá me ayudás?. piensa está enojado. porque no le ayudó nada. su mamá no le ayudó nada. sí. los nombres de su hermana. tiene dos hermanos éste. porque quería escribir una carta y no le salió nada. se fija que ahora está contento porque le salieron las letras.	17
0,566	2	está pensando qué puede escribir. y puede escribir una carta, eh... otra vez lo mismo!. que va a hacer. sí, escribiendo. y que piensa, que está haciendo lo que está pensando y mientras... y porque vio que le salió todo mal, le salió rayas y todo eso. sí, y lo hace otra vez. ahora lo pone en la carta y se lo manda, al cartero. y para ver si le salió bien. ahí la rompe. acá parece que no. bueno... ya terminé... sí... claro...	8
0,572	3	está pensando en qué puede hacer. una carta. en escribir algo lindo. en escribir un cuento. le habrá salido un poco mal lo del cuento. porque le está saliendo muy mal. porque por ahí está pensando en hacer algo lindo y no lo puede hacer y lo tiene que borrar. eh, no sé. sí. en las letras que están bien. en qué ponerla. sí, en un sobre. acá. que le ayuden a escribir la carta que quiere escribir. no, que le ayuden a hacer la letra.	20

Tabla 7. Respuestas típicas de la modalidad "preescolar".

Groupe d'individus : G1

Critère de classement	Numéro	Libellé de la réponse	Identificateur
0,553	1	en escribir palabras o dibujitos en hacer una batata. y en escribir el nombre de la batata. ella tiene que pensar porque su mamá le puede decir como escribe batata y tiene que escribir con bu, con la be. y tiene que pensar escribir oso y eso es muy cortito y su mamá le tiene que decir que empieza con la o y empieza con la o y termina con la o. lo tiene que escribir. que cuando ya terminó de escribir tiene que decirle a su mamá que juegue con barre a hacer un muñeco. está pensando en escribir un moño. porque no le gustó lo que hizo. un muñeco de nieve. y puede escribir el dos mil porque hace falta. porque ya pasó el dos mil. en hacer un automóvil.	27
0,583	2	que la nena?. qué?. que tienen que escribir la tarea y que ella también tiene que ayudarle a los hermanos cuántos años?. porque yo tengo seis años y ella debe tener seis porque mi mamá me enseñó mucho, mucho a hablar y no decirle malas palabras al papá. porque el papá después se enoja y le va a pegar y si yo le digo malas palabras a mi papá y mi papá me pega, a mí no me gusta que mi papá me pegue, entonces si a mí no me gusta que mi papá me pegue yo no tengo que decir malas palabras. que, que ella, que ella, terminó de hacer la tarea y va a salir y la mamá le va a revisar todo y la mamá después cuando viene no la deja salir más. el timbre?. después voy. porque hizo la tarea mal, hizo todo mal. porque... se dio cuenta que... que estaba todo mal, por. por eso, sino no lo iba a borrar. y cuenta el cuentito. a dónde?. no sé. no sé.	39
0,583	3	está pensando, primero del día va a pensar porque siempre el día está al empezar. que dibuje una foto viste como de los álbumes dibujar una foto. que está escribiendo nadia. el apellido nadia micaela zuber. me parece que está borrando porque le salió mal la letra de nadia. chi, chi, chi borra y escribe de vuelta. porque estaba mirando para la pared. estaba mirando el dibujito de los días que hay la casita ahí. no sé, qué se yo. en el papel dice nadia me cae la zuber te quiero invitar a mi fiesta de cumpleaños de quince por favor vení te mando la tarjeta y el mensaje. lo está doblando para que lo manden se vaya a su casa y lo manden, el mensaje y sale el papel de su casa y va para allá.	28

Tabla 8. Respuestas típicas de la modalidad "primer grado".

Groupe d'individus : G4

Critère de classement	Numéro	Libellé de la réponse	Identificateur
0,486	1	las cosas que tiene que escribir en la carta, porque si uno no piensa las cosas que tiene que escribir en la carta no puede escribir. las cosas que tiene que seguir para escribir, porque hay algunos que escriben hasta la mitad de la carta y después tienen que seguir para abajo, y tiene que pensar que va en las cosas que tienen que seguir para abajo. porque sino no la pueden escribir completa. porque algo le salió mal. porque hay algunas cartas que le quieren decir un montón de cosas y la letra le puede salir mal y si lo escribe con lapicera no lo puede borrar o con lápiz si se puede borrar porque con la goma se borra. porque al llegar al final de la carta lo lee todo entero y si tiene algún error lo borra con la goma o con alguna otra cosa. está pensando si la van a recibir al que está mandando que por ahí alguno que no lo recibe. sí. para que corrija los errores que tiene. sí. y alguna letra, alguna letra que escribió la carta y la letra está mal, o se olvidó de poner alguna palabra. que escribirle sí y si no sabe el nombre de su tío o de su hermano o esa cosas le va a preguntar a su mamá, no sé si le voy a preguntar a mi mamá o no. qué decirle a su tío o a su tía. sí cosas así como están mis primos cómo está mi hermano, cómo están mis sobrinos. algo le habrá salido mal. alguna palabra que ella no sabe. porque ella mientras que escribió toda la carta después la repasó y vio que se equivocó en una letra y ahí la borró. que los extraña mucho, mientras que va leyendo va pensando en sus tíos y en sus primos. alguna palabra que le sale mal o algún dibujo que le haya hecho. que se van a poner contentos sus tíos.	100
0,486	2	y, me imagino que estar pensando, una historia, tiene que pensar algo que tenga sentido, una oración, algo, como que el chico está jugando en la plaza. capaz, que está pensando como se escribe una letra, o si helado va con hache o no, o pensando como puede armar una oración, como puede terminar el cuento.. capaz que borra porque le salió mal una letra, o una palabra y escribió cualquier cosa. y, pasaría que está escribiendo, y en un momento escribe por ejemplo: la tía de este chico está enferma? y, ahí se acuerda que su abuela está enferma, y se pone a pensar sobre eso, y no lee lo que había escrito, y se pone a contar otra historia. capaz que se fija si tiene alguna falta de ortografía, o, si se voló y después siguió escribiendo y no se dio cuenta.	93
0,509	3	y, me imagino que estar pensando, una historia, tiene que pensar algo que tenga sentido, una oración, algo, como que el chico está jugando en la plaza. capaz, que está pensando como se escribe una letra, o si helado va con hache o no, o pensando como puede armar una oración, como puede terminar el cuento.. capaz que borra porque le salió mal una letra, o una palabra y escribió cualquier cosa. y, pasaría que está escribiendo, y en un momento escribe por ejemplo: la tía de este chico está enferma? y, ahí se acuerda que su abuela está enferma, y se pone a pensar sobre eso, y no lee lo que había escrito, y se pone a contar otra historia. capaz que se fija si tiene alguna falta de ortografía, o, si se voló y después siguió escribiendo y no se dio cuenta.	84

Tabla 9. Respuestas típicas de la modalidad "cuarto grado".

Los resultados del procedimiento CORDA se muestran en la Tabla 10 (para la palabra: cuento) y en la Tabla 11 (para la palabra: cómo). Este procedimiento permite situar cada palabra en su contexto original de aparición. Los números que se encuentran en la tercera columna permiten identificar al niño que proporciona el fragmento de esa

respuesta y por tanto al grupo en que se encuentra. En nuestro caso los niños 1 a 20 son de preescolar, 21 a 40 de primer grado y 81 a 100 son de cuarto grado. Las palabras solicitadas en este caso fueron: cuento y cómo.

Contextes du mot: cuento

está pensando que estudia un	cuento		19
en escribir un	cuento		20
le habrá salido un poco mal lo de	cuento		20
está pensando un	cuento		23
el	cuento		24
los dibujos y cómo era e	cuento	toda la historia	24
tramando un	cuento	de terror o bonito	26
pensar que cómo haber supónete que el quiere escribir	cuento	y él y él va repitiendo muchas veces a ver cuál es la que	26
estar pensando que	cuento	más lindo que la flor	26
pensando qué escribir para hacer un	cuento		31
pensando en escribir que este	cuento	de un perro y y escribir muchas cosas para que se haga un	31
de un perro y y escribir muchas cosas para que se haga un	cuento		31
y ellos están contando un	cuento	de un payaso ahí está	31
en escribir un	cuento	del payaso y después escribir lo que pensó para escribir un	31
del payaso y después escribir lo que pensó para escribir un	cuento		31
un	cuento		32
y si escribe un	cuento	estar pensando en el final del cuento o en el principio y si	81
y si escribe un cuento estar pensando en el final del	cuento	o en el principio y si escribe una carta estar pensando en	81
en e	cuento		82
pensando como puede amar una oración como puede terminar e	cuento		84
cómo sigue e	cuento		86
cómo puede empezar e	cuento	de qué se puede tratar y todo eso	90
un	cuento		91
que quiere poner una carta a su mamá un	cuento	a su papá	97

Tabla 10. Contexto léxico de la palabra "cuento".

Contextes du mot: cómo

	cómo	se llama para leer	13
	cómo	escribir	14
	cómo	hacer las letras	14
	cómo	se escribe	22
los dibujos y	cómo	era el cuento toda la historia	24
pensar que	cómo	haber suponéte que el quiere escribir cuento y él y él va	26
	cómo	se escribe eso cómo se escribe	38
cómo se escribe eso	cómo	se escribe	38
	cómo	se escribe algo o qué les puede escribir a sus parientes	83
	cómo	va a escribirlo cómo va a empezar	86
cómo va a escribirlo	cómo	va a empezar	86
	cómo	sigue el cuento	86
te ocurre algo lo escribís después tenés que seguir pensando	cómo	sigue	86
preguntándole a alguien o acordándose de	cómo	era la palabra	86
otro es lo mismo que yo empiezo a escribir y de pronto no sé	cómo	seguir	87
	cómo	empezar	89
	cómo	puede empezar el cuento de qué se puede tratar y todo eso	90
	cómo	escribir	92
en	cómo	terminar	92
sí cosas así como están mis primos	cómo	está mi hermano cómo están mis sobrinos	93
sí cosas así como están mis primos cómo está mi hermano	cómo	están mis sobrinos	93
	cómo	puede escribir la carta	95
	cómo	recibir la carta si se equivocó	96
va a poner que	cómo	es que los quiere que les manda besos cómo es y que va a	98
va a poner que cómo es que los quiere que les manda besos	cómo	es y que va a	98

Tabla 11. Contexto léxico de la palabra "cómo".

Apéndice: Algunas aclaraciones

La mayoría de las aclaraciones que se presentan a continuación fueron extraídas del texto entregado en el curso “SPAD.T Técnicas cuantitativas de análisis textual en el estudio de representaciones sociales”, impartido por Valencia y Elejabarrieta en la Universidad de Buenos Aires, en 1995.

Umbral (en procedimiento MOTs). Hemos indicado que el umbral establece un corte en el corpus que es necesario para el análisis. La determinación de este punto es un tema delicado y especial. No hay criterios estándares para determinar cuál es el punto que separa el discurso colectivo sobre un objeto, del discurso individual sobre el mismo. Depende en cada caso de la naturaleza del discurso, del tamaño del colectivo, del objeto referenciado, etc. Podemos sin embargo plantear tres criterios: el primero es que, sea como sea, para decir que estamos ante un discurso colectivo, cuantitativamente, debemos fijar un umbral de repetitividad mínimo. El segundo es que cuanto mayor sea el umbral, aparentemente de mejor modo estaremos dando cuenta de la regularidad colectiva. Sin embargo, esta regularidad puede ser engañosa en cuanto a su estructuración; por ejemplo cuando todos los entrevistados usan una de las palabras que

está presente en una de las preguntas, cuando usan una "muletilla" como, este..., este..., etc.. La regularidad excesiva puede proporcionar truismos lingüísticos y hacer perder partes interesantes de los discursos grupales. El tercero es que dado un corpus textual o conjunto de textos, el umbral de representatividad de una forma lingüística debe fijarse en base a los subgrupos que componen el colectivo analizado y no en base al total de la muestra. De lo contrario podemos perder la representatividad grupal y las diferencias intergrupales en la conformación discursiva. Estas cuestiones son consideradas importantes en el estudio de las concepciones o representaciones grupales. Si bien no hay criterios estándares para la fijación del umbral, es preciso señalar que en una tesis (Crego, 1998) se propuso el concepto de umbral óptimo. El umbral óptimo sería el que permite la mejor discriminación de los grupos para cada análisis. Se obtiene probando distintos umbrales y analizando los resultados que se obtienen en cada caso.

La solución factorial en el análisis de correspondencias. En este análisis la solución factorial, a diferencia de lo que sucede con las variables cuantitativas, no tiene excesiva importancia. Puesto que todos los índices están en relación con el número de filas/columnas, cuando se trata de un procesamiento sobre una tabla léxica, la solución suele ser en general mala, generalmente la cantidad de palabras distintas, de columnas es muy grande, considerándose aceptable una explicación de varianza total superior al 15%. El caso de una tabla léxica agregada, es inverso. Puesto que la cantidad de filas/columnas es pequeña, la solución factorial suele ser bondadosa para la sumatoria de varianza explicada de los factores retenidos. En este caso se considera aceptable un conjunto de explicación de alrededor del 75%.

Para decidir cuántos ejes deben conservarse, hay más de una respuesta. La primera, aplicable a situaciones en las que el número total de ejes sea bajo, es: tantos como sea posible. Un segundo criterio, aplicable principalmente a situaciones en las que el número de ejes es elevado, es: los ejes que se encuentran por encima de un decrecimiento brusco (o salto) en el porcentaje de inercia explicado. Un tercer criterio es el del valor propio medio, es decir los ejes asociados a los valores propios que sean mayores que la media.

La interpretación de un análisis de correspondencias se realiza en base a contribuciones absolutas y, gráficamente, en las posiciones de las coordenadas. Hay dos sistemas para determinar qué valores deben ser considerados válidos como contribuciones absolutas, es decir como contribución de una fila o como conformación de una columna a la

conformación de un factor. Para ello se divide el 100% de la contribución del factor por el número de elementos intervinientes, sean filas o columnas. Cuanto más alta es la contribución, tanto más importante es la fila o la columna en la formación del factor. No obstante, en el análisis factorial, las contribuciones muy altas (por ejemplo, del 20%) suelen ser mal recibidas porque concentran en un factor punto fila o columna y factorializan poca cosa. Cuando se obtiene un resultado de este tipo, lo más aconsejable es repetir el análisis desechando el punto fila o columna que satura excesivamente el factor.

La *interpretación* se realiza en los gráficos de cruce de los ejes de coordenadas. Las posiciones en los ejes fila/columna guardan relación con las contribuciones absolutas. La interpretación es más rica en las gráficas por los cruces de los ejes. En la interpretación se suelen considerar:

- a) Los cuadrantes
- b) Las asociaciones alejadas del centro
- c) La representatividad de las cadenas alejadas en relación al corpus total original.

Coordenadas factoriales. Son la posición métrica de cada fila o columna en el plano de un eje factorial

Valores test. Son los valores que determinan si la posición de una categoría sobre un eje es significativamente diferente respecto al resto de las categorías de la variable considerada. El valor test se interpreta de forma similar a la t de Student (el cálculo es semejante). Se considera valor test significativo, es decir categorías posicionadas significativamente, diferentes del resto, a los valores superiores a 1,96 para una probabilidad de $<.05$.

Distancia χ^2 . En el AFC la semejanza entre dos filas está definida por una distancia *ponderada* según la importancia de las columnas. Esta distancia se la conoce con el nombre de distancia del Ji o Chi Cuadrado (χ^2). El cuadrado de esta distancia es una suma ponderada de los cuadrados de las diferencias término a término entre los dos perfiles, es decir el cuadrado de una distancia euclídea ponderada entre perfiles.

Remitimos a Baccalá y Montoro (2008) para una formulación matemática de la misma. Las autoras observan que esta distancia es adecuada pues: a) si los perfiles fuesen idénticos esta distancia sería cero y cuanto más alejados estén esta distancia es mayor y b) en la diferencia entre dos perfiles filas, una modalidad será más contributiva a la comparación cuanto más raro sea el atributo que presenta. Una situación análoga se registra para las columnas.

Coordenadas. Permiten identificar las posiciones reales de las categorías. Su inserción en el gráfico de análisis permite establecer el sentido de las diferencias indicadas por los valores test.

Agradecimientos

Agradecemos las cuidadosas observaciones y sugerencias de índole matemática y estadística realizadas por un revisor.

Referencias

De los métodos

Baccalá, N. y Montoro, V. (2008). *Introducción al Análisis Multivariado*. Cuaderno Universitario n° 51. Centro Regional Universitario Bariloche. Universidad Nacional del Comahue.

Bécue, M. (1991). *Análisis estadístico de datos textuales: métodos de análisis y algoritmos*. París: CISIA.

Benzécri, J. (1973). *L'Analyse des Données*. París: Dunod.

Crego, C. (1998). *Indagando sobre los umbrales*. Tesina de licenciatura no publicada. Universidad Nacional del Comahue, San Carlos de Bariloche, Argentina.

Crivisqui, E. M. (1993). *Análisis Factorial de Correspondencias*. Asunción: Universidad Católica.

Decisia (2002). Spad rel. 5.5: Système pour l'analyse des données. Logiciel diffusé par Decisia, Levallois-Perret, Francia.

de la Cruz, M. Apuntes no publicados elaborados a partir del curso "SPAD.T Técnicas cuantitativas de análisis textual en el estudio de representaciones sociales", dictado por

Profesores J. Valencia. y F. Elejabarrieta y organizado por la Universidad de Buenos Aires. Buenos Aires, 1995.

de Torres Curth, M. (2005). *Tomar apuntes en la universidad*. Tesis de maestría no publicada. Universidad Nacional del Comahue, Neuquén, Argentina.

García Calvo (1991). Prólogo en M. Bécue, *Análisis estadístico de datos textuales: métodos de análisis y algoritmos*. París: CISIA.

Germano, A. (2002). *Una aplicación de las técnicas básicas del análisis multivariado al análisis de datos provenientes de una entrevista que indaga conocimientos numéricos en niños*. Tesis de licenciatura no publicada. Universidad Nacional del Comahue, San Carlos de Bariloche, Argentina.

Lebart, L. (1982). L'analyse statistique des réponses libres dans les enquêtes socio-économiques, *Consommation - Revue de Socio-Economie*, (1), 39-62.

Lebart, L. y Salem, A. (1994). *Statistique textuelle*. Paris: Dunod.

Lebart, L., Salem, A. y Bécue Bertaut, M. (2000). *Análisis estadístico de textos*. Lleida: Milenio.

Montoro, V. (2003). *Estudio exploratorio sobre concepciones de los alumnos ingresantes a la universidad respecto de la noción de infinito matemático*. Tesis de maestría no publicada. Universidad Nacional del Comahue, Neuquén, Argentina.

De las aplicaciones utilizadas como ejemplo

Baccalá, N., de la Cruz, M. y Scheuer, N. (2002). Una aplicación de la lexicometría a la descripción de procesos evolutivos en psicología *JADT 2002. 6èmes Journées internationales d'Analyse statistique des Données Textuelles* (pp. 77-87). Saint Maló: IRISA-INRIA. Vol. 1.

de la Cruz, M., Scheuer, N., Huarte, M. F., Sola, G. (2002). La mirada de padres y maestros sobre qué ocurre en la mente del niño cuando escribe. *Educación y Pedagogía*, (XIII) 31, 175-185.

Scheuer, N., de la Cruz, M., Pozo, J. I., Huarte, M. F., y Sola, G. (2006). The mind is not a black box: Children's ideas about the writing process. *Learning and Instruction*, 16 (1), 72-85.

Scheuer, N., de la Cruz, M. y Pozo, J. I. (2010). *Aprender a dibujar y a escribir. Las perspectivas de los niños, sus familias y maestros*. Buenos Aires: Noveduc.