



*Las opiniones y los contenidos de los trabajos publicados son responsabilidad de los autores, por tanto, no necesariamente coinciden con los de la Red Internacional de Investigadores en Competitividad.*



Esta obra por la Red Internacional de Investigadores en Competitividad se encuentra bajo una Licencia Creative Commons Atribución-NoComercial-SinDerivadas 3.0 Unported. Basada en una obra en riico.net.

## Uso de software lexical: una revisión comparativa

*Cecilia Morales del Río<sup>1</sup>*

### **Resumen**

Las investigaciones cualitativas en las ciencias sociales se están apoyando más y más en el uso de software como apoyo a las investigaciones cualitativas. El seleccionar el software adecuado es el primer obstáculo que se enfrentan los investigadores, entre los que existen en el mercado: de pago o gratuitas, es difícil discernir y escoger de una primera opción. Esta investigación aborda un análisis de cuatro de estas herramientas y sus aplicaciones, así como un estudio comparativo de sus módulos o secciones (T-LAB, ALCESTE, IRAMUTEQ y VOYANT-TOOLS). En inicio, un análisis de literatura para ver sus ventajas y desventajas. Como segundo lugar se utilizaron los cuatro con la misma base de datos y se realizaron los mismos análisis en forma subjetiva para evaluar el desempeño de los mismo. Y, por último, utilizando el método Delphi pusimos a consideración de un panel de expertos que evaluaron y valoraron los software, en el ámbito de una investigación cualitativa.

**Palabras clave:** Software de análisis lexical, ALCESTE, T-LAB, IRAMUTEQ, VOYANT-TOOLS, Comparativa de software, Análisis cualitativo en administración.

### **Abstract**

In recent years, has been emerged a new generation software for the lexical analysis, as support to qualitative research. This research focuses on create a comparative analysis on applications and modules software (T-LAB, ALCESTE, IRAMUTEQ and VOYANT-TOOLS). Starts with a literature analysis, with the most representative, in order to discover advantages and disadvantages. Furthermore, had been used with the same database, running the same analysis to evaluate their performance. Finally, a panel of experts in qualitative research, assessed and evaluated with Delphi method, those softwares. This study highlights the importance to discern which are the potentials within perceptions in use on the qualitative analysis on management sciences.

**Keywords:** Lexical analysis software, ALCESTE, T-LAB, IRAMUTEQ, VOYANT-TOOLS, software testing, management qualitative analysis.

**JEL: M15, O35**

---

<sup>1</sup> Universidad de Guadalajara, Centro de Ciencias Económico Administrativas

## **Introducción**

Es difícil hoy en día olvidar la contribución del software de análisis lexical que son extremadamente potentes. Estas herramientas son una ayuda muy eficaz y garantizan la sistematicidad del análisis (Cortini & Tria, 2014). Permiten analizar las intuiciones que uno puede tener a simple vista del corpus (Kuz, Falco, Nahuel, & Giandini, 2015), además de destacar algunos de los aspectos que un análisis manualmente no revelaría. Son una ayuda para la lectura no secuencial de texto, lo cual es útil para diferentes tipos de análisis (de Alba, 2004).

Universidades y/o grupos de investigación han diseñado algunos software para el ámbito del estudio de Análisis del Contenido (Reinert, 1986), Análisis semántico (Gambetti & Graffigna, 2010), Análisis temático, Minería de textos y más. Algunos de ellos especializados para el análisis lexical: ALCESTE, T-LAB, VOYANT-TOOLS, IRAMUTEQ, FeatureLens, Lexico3, DtmVic, AntConc, CasualConc, SCP y WordSmith Tools.

Algunos investigadores cualitativos Reinert (2003), Peyrat-Guillard (2006), Delavigne (2014), Kalampalikis & Moscovici (2005), Kalampalikis, Abric, & Kalampalikis (2003), Einert (1983), Azevedo, Costa, & Miranda (2013), La société IMAGE (2015), Cortini & Tria (2014), Lancia (2017), Crook, Todd, Combs, Woehr, & Ketchen Jr. (2011), trabajan utilizando estos software de análisis lexical por su capacidad para generar las co-ocurrencias pero en su mayoría siguen utilizando los sistemas clásicos de análisis cualitativo: Atlas.ti, Nvivo, SPSS Modeler, u otros. Por lo que se cree necesario el realizar este tipo de comparativos para que sean más tomados en cuenta para facilitar el trabajo de los investigadores.

El documento presenta un análisis comparativo de cuatro de estos software, los principales, que de acuerdo con los expertos son los más utilizados: ALCESTE, IRAMUTEQ, T-LAB y VOYANT-TOOLS (Álvarez Cadavid, Giraldo Ramírez, & Navarro Plazas, 2017), en primer lugar analizando bibliográficamente lo que se ha escrito y publicado de ellos en el ámbito científico, describiendo las principales características de cada uno. En un segundo lugar, se utiliza la misma base de datos para conocer de primera mano los diferentes tipos de gráficas, de usabilidad, de desempeño y conveniencia de uso de cada uno de ellos. Finalmente, se ha convocado a expertos en el uso de los mismos a contestar un cuestionario para que califiquen cada uno de ellos, para así poder realizar un comparativo.

Todo esto con el objetivo de permitir a los investigadores cualitativos hacer más fácil el poder determinar cuál es indicado para cada análisis de texto que cada investigador requiere para su proyecto de investigación.

Este documento contiene cinco apartados principales: una introducción al tema del análisis lexical y su contexto; el compendio de los resultados del análisis bibliográfico, de los artículos encontrados hasta 2019 más representativos de investigaciones con el uso de estos cuatro softwares; la evaluación subjetiva en un

contexto dinámico-práctico; los resultados del panel de expertos, y finalmente las conclusiones y recomendaciones.

### **Metodología**

La metodología a utilizar es la Triangulación de datos, de acuerdo a Hernández Sampieri (2010), al hecho de utilizar diferentes fuentes y métodos de recolección, se le denomina triangulación de datos. Esta triangulación consta de tres fuentes y métodos:

- 1) En primer lugar, un análisis de literatura de los cuatro software. “Una fuente muy valiosa de datos cualitativos son los documentos, materiales y artefactos diversos, que nos pueden ayudar a entender el fenómeno central de estudio” (Hernández, Fernández, & Baptista, 2010).
- 2) Se sigue de una disertación de sus uso y alcances. Se analiza el mismo corpus de datos con el eWOM (Word-of-Mouth electrónico) de un hotel en Puerto Vallarta con los datos de 30 huéspedes del 2019. Con esta base se realizaron el análisis en los cuatro software, para analizar la usabilidad, compatibilidad con otros software, la facilidad de uso, el rendimiento general del mismo.
- 3) Y finalmente un Focus group a un panel de expertos con el método Delphi.

### **Implicaciones sobre la importancia de los software lexicales**

El uso de software para el análisis de datos ya no es algo nuevo, más sin embargo, en una revisión realizada en 2017, los investigaciones indagaron por el uso específico del análisis mediante software especializado y encontraron que apenas el 4% de los investigadores consultados lo utilizaba (Álvarez Cadavid et al., 2017).

La línea divisoria entre el análisis cuantitativo y el análisis cualitativo de textos (CAQDAS: *Computer-Aided Qualitative Data Analysis Software*) es actualmente la intervención humana. Es necesaria, pero la intervención humana realiza una serie de tareas previas al análisis en sí, mientras que el análisis es obra humana, aunque pueda ser asistido mediante software (Colina, Pericás, Miquel, Olivé, & Roldán, 2003). El avance en el reconocimiento del lenguaje en forma natural, llevará a una casi completa automatización de estos análisis (Stefanello, Francisco, & Carranza, 2010).

Dentro del análisis cualitativo y de redes sociales, el análisis de contenido y en específico el análisis de textos han evolucionado día con día, hasta elaborar herramientas sofisticadas para tal objetivo (Leonidas Aguirre, 2011).

De acuerdo a la Real Academia Española, la palabra **lexical** es relativo o perteneciente a *léxico*, que expresa contenidos o conceptos propios del léxico, por oposición a los característicos de la gramática. Y además de un conjunto de palabras de un idioma, o de las que pertenecen al uso de una región, a una actividad determinada, a un campo semántico dado, etc. (RAE, 2017). Por lo que el vocablo *lexical*

tendría más relación con la semántica léxica, que es la rama de la semántica que estudia el significado de las palabras, así como las diversas relaciones de sentido que se establecen entre ellas.

El análisis de datos lexicales o las estadísticas de texto es la metodología que se propone descubrir la información esencial contenida en un texto (Huisman & van Duijn, 2004). Esta metodología, en un fuerte crecimiento, encuentra sus aplicaciones en todas las áreas en donde se necesita de secuencias de texto para extraer el significado o representación (La société IMAGE, 2015). Más allá de la sola dimensión del léxico, o la semiótica, el objetivo, por lo tanto, es también semántico. Las unidades lexicales son palabras, simples o múltiples, archivadas y clasificadas en base a un cierto criterio (Lancia, 2017).

En el mundo de los textos, existe el grupo de palabras que generan un fragmento en el discurso. El análisis lexical más utilizado o famoso es el que utiliza el método de Reinert, el cual consiste en la idea de que todos los discursos son expresados en grupos lexicales que se pueden organizar racionalmente y dar coherencia a las partes que enuncian. Mediante el algoritmo Reinert se analizan las situaciones, pensamientos, emociones en los mundos lexicales o representaciones sociales que se comparten por los internautas (Larruzea-Urkixo, Cardeñoso Ramírez, & Idoiaga Mondragon, 2019).

Dentro del lenguaje de los software lexicales, encontraremos su propia narrativa y sus términos específicos, así que nos dimos a la tarea de explorar algunos y detallarlos para estar en sintonía con los términos utilizados entre ellos. Los principales son: *corpus*, *cluster*, *collocation*, *concordance*, *keyword* y *n-grama* (Peyrat-Guillard, 2006).

**Corpus.** La palabra corpus es latina para el cuerpo (plural *corpora*) (RAE, 2017). En la lingüística un corpus es una colección de textos ('cuerpo' del lenguaje) que se almacenan en una base de datos electrónica (Baker, Hardie, & McEnery, 2006).

**Cluster.** Un término usado para describir cualquier grupo de palabras en secuencia. Un conjunto de textos que estadísticamente tiene las características lingüísticas similares (Baker et al., 2006).

**Colocación (Collocation).** La colocación es el fenómeno que rodea el hecho de que ciertas palabras son más probables de ocurrir en combinación con otras palabras, en ciertos contextos. Una colocación es por lo tanto una palabra que se produce dentro de la vecindad de otra palabra (Baker et al., 2006).

**Concordancia (Concordance).** Una concordancia es una lista de todas las apariciones de un determinado término de búsqueda en un corpus, que se presentan en el contexto en que se producen. Un término de búsqueda es a menudo una sola palabra a pesar de que muchos de concordancia de los programas permiten a los usuarios buscar en el *clúster* de las frases, las palabras que contienen los comodines, las etiquetas o las combinaciones de palabras y etiquetas. Las concordancias generalmente puede ser ordenada alfabéticamente por el término de búsqueda en sí o a X lugares a la izquierda o a la derecha del término de búsqueda, permitiendo a los patrones lingüísticos para ser más fácilmente observado por los seres humanos (Baker et al., 2006).

**Palabra clave (Keyword).** Una palabra que aparece en un texto o corpus de forma estadísticamente significativa con más frecuencia de lo que cabría esperar por azar cuando se compara con un cuerpo que está a mayor o de igual tamaño (Azevedo et al., 2013).

**N-grama.** Una secuencia de n letras de una cadena de caracteres después de la eliminación de espacios en blanco (Baker et al., 2006).

Una vez explicados los términos más comúnmente utilizados por los software de análisis lexical, se realizó una investigación en la *world wide web* sobre las posibilidades y opciones disponibles actualmente, enfocándose principalmente en los realizados por investigadores cualitativos. Se tomaron en cuenta los puntos como creadores, disponibilidad de plataforma y documentación disponible, como manuales. Los resultados se muestran en la Tabla 1.

**Tabla 1.** Principales software utilizados actualmente en la *world wide web*.

| Software                        | Última versión disponible | Creadores   | Disponibilidad  | Manuales  | Link software   |
|---------------------------------|---------------------------|---|---|---|---|
| Alceste                         | v. 8, 2018                | Max Reinert, The Image society, National Center for Scientific Research (CNRS)  | Windows   | <a href="http://www.image-zafar.com/APK/DOC2015.zip">http://www.image-zafar.com/APK/DOC2015.zip</a>   | <a href="http://www.image-zafar.com/Logiciel.html">http://www.image-zafar.com/Logiciel.html</a>                 |
| AntConc                         | v. 3.4.4, 2014            | Laurence Anthony, Center for English Language Education (CELESE), Faculty of Science and Engineering, Waseda University | Windows, Macintosh OS X, y Linux.                     | <a href="http://www.laurenceanthony.net/software/antconc/releases/AntConc344/help.pdf">http://www.laurenceanthony.net/software/antconc/releases/AntConc344/help.pdf</a> | <a href="http://www.laurenceanthony.net/software/antconc/">http://www.laurenceanthony.net/software/antconc/</a> |
| CasualConc                      | v. 2.0.6, 2016            | Yasu Imao, Osaka University   | Mac OS X  | <a href="https://sites.google.com/site/casualconc/download">https://sites.google.com/site/casualconc/download</a>   | <a href="https://sites.google.com/site/casualconc/Home">https://sites.google.com/site/casualconc/Home</a>       |
| DtmVic                          | v. 6.0, 2018              | L. Lebart, A. Morineau ley 1901: CESIA  | Windows XP, Vista, Windows 7, Windows 8.1, Windows 10 | <a href="http://www.dtmvic.com/doc/DtmVic_English_Manual_2016.pdf">http://www.dtmvic.com/doc/DtmVic_English_Manual_2016.pdf</a>   | <a href="http://www.dtmvic.com/05_SoftwareE.html">http://www.dtmvic.com/05_SoftwareE.html</a>                   |
| FeatureLens                     | 2007                      | Anthony Don University of Maryland  | MacOSX, Linux and WinXP                               | <a href="http://monkproject.org/downloads/featurelens/doc/featureLensSystem.pdf">http://monkproject.org/downloads/featurelens/doc/featureLensSystem.pdf</a>             | <a href="http://www.cs.umd.edu/hcil/textvis/featurelens/">http://www.cs.umd.edu/hcil/textvis/featurelens/</a>   |
| IraMuTeQ                        | v. 0.7 alpha 2            | Pierre Ratinaud Sébastien Déjean de la Université de Toulouse. David Skalinder de Mash Strategy                         | Windows, Macintosh OS X, y Linux.                     | <a href="http://www.iramuteq.org/documentation/fichiers/guia-iramuteq">http://www.iramuteq.org/documentation/fichiers/guia-iramuteq</a>                                 | <a href="http://www.iramuteq.org">http://www.iramuteq.org</a>   |
| Lexico3                         | v. 5.8.1, 2019            | Lexi&Co   | Windows   | <a href="http://lexico.com/ressources/L3-usermanual.pdf">http://lexico.com/ressources/L3-usermanual.pdf</a>   | <a href="http://lexico.com/">http://lexico.com/</a>   |
| SCP, Simple Concordance Program | v. 4.0.9, 2016            | Alan Reed   | Windows y Macintosh OS X                              | <a href="http://www.textworld.com/scp/scpPics/tour/index.html">http://www.textworld.com/scp/scpPics/tour/index.html</a>   | <a href="http://www.textworld.com/scp/index.html">http://www.textworld.com/scp/index.html</a>                   |

|                 |              |   |                                     |   |   |
|-----------------|--------------|---|-------------------------------------|---|---|
| T-lab           | v. Pro 2019  | Franco Lancia                                 | Windows 7, 8, 10                    | <a href="http://www.mytlab.com/Manual_es_plus.zip">http://www.mytlab.com/Manual_es_plus.zip</a>   | <a href="http://tlab.it/es/presentation.php">http://tlab.it/es/presentation.php</a>   |
| Voyant-tools    | v. 2.2, 2017 | Stéfan Sinclair y Geoffrey Rockwell           | Online                              | <a href="http://docs.VOYANT-TOOLS.org/about/">http://docs.VOYANT-TOOLS.org/about/</a>   | <a href="http://VOYANT-TOOLS.org">http://VOYANT-TOOLS.org</a>                         |
| WordSmith Tools | v. 7, 2016   | Mike Scott, Lexical Analysis Software, Stroud | Windows XP<br>Windows 7, 8, 8.1, 10 | <a href="http://lexically.net/downloads/version7/HTML/index.html?getting_started.htm">http://lexically.net/downloads/version7/HTML/index.html?getting_started.htm</a> | <a href="http://www.lexically.net/wordsmith/">http://www.lexically.net/wordsmith/</a> |

Fuente: Elaboración propia.

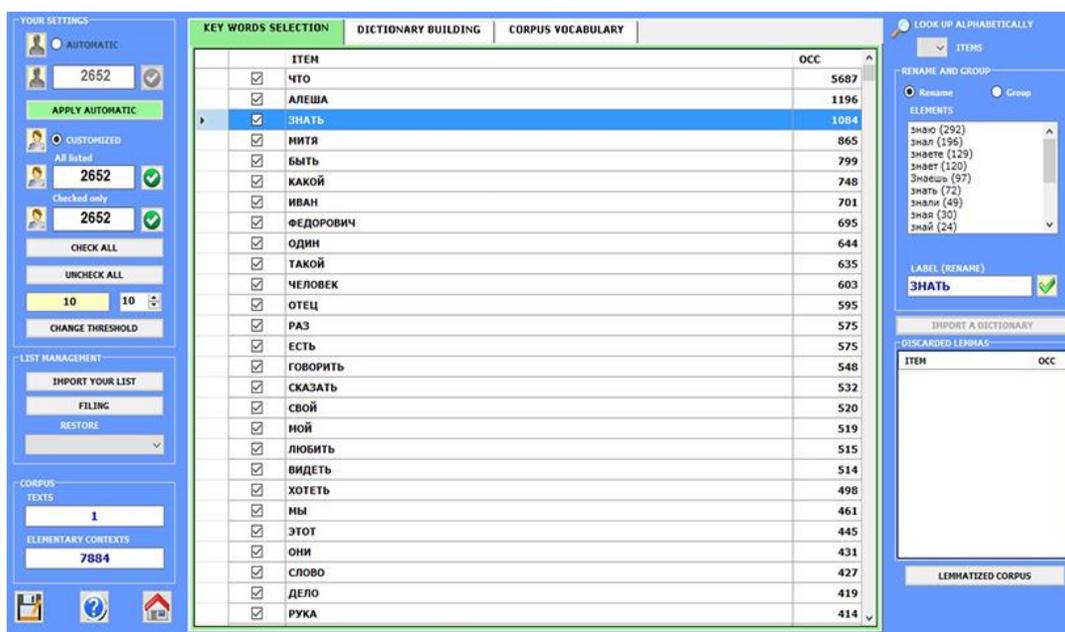
Después de este análisis general, se ha decidido solamente generar la comparativa con ALCESTE, IRAMUTEQ, T-LAB y VOYANT-TOOLS, siendo los que han seguido actualizándose hasta la fecha, y los que al parecer han ganado más reputación en el ambiente de los investigadores.

### **Resultados de análisis bibliográfico de los software T-LAB.**

T-LAB es un software de análisis cuantitativo compuesto por un conjunto de herramientas lingüísticas y estadísticas. Diseñado para plataforma Windows por Franco Lancia y se encuentra en su versión 7.1. (Stefanello et al., 2010). Cuenta con varias herramientas las cuales permiten a los investigadores los siguientes análisis:

- Explorar, medir y mapear las relaciones de concordancia entre palabras-clave;
- Personalizar y aplicar, tanto al análisis lexical como al análisis de contenido, diferentes tipos de diccionarios;
- Clasificación automática de las unidades de contexto y de los documentos, bien a través de una metodología *top-down* o bien utilizando una metodología *bottom-up*;
- Aplicar categorías para la *sentiment analysis*;
- Comprobar qué unidades lexicales, qué unidades de contexto y qué temas son 'típicos' de subconjuntos específicos de determinados textos;
- Generar mapas semánticos que representen aspectos dinámicos del discurso;
- Ejecutar diferentes tipos de análisis de las correspondencias y de análisis de los clústeres;
- Analizar todo el corpus o sólo algunos de sus subconjuntos utilizando diferentes listas de palabras-clave.
- Verificar los contextos de ocurrencia y/o concordancias de palabras y lemas;
- Crear, explorar y exportar diferentes tipos de tablas de contingencia y matrices de concordancias (Lancia, 2017).

**Figura 1.** Interface con las herramientas de T-Lab.



Las unidades de análisis de T-LAB son de dos tipos: unidades lexicales y unidades de contexto. Las unidades lexicales son palabras, simples o múltiples, archivadas y clasificadas en base a un cierto criterio. En la base de datos T-LAB cada unidad lexical constituye un registro clasificado con dos campos: palabra y lema (Gambetti & Graffigna, 2010). En “palabra” se cuentan las palabras como aparecen en el corpus, en el “lema”, se enumeran las etiquetas atribuidas a grupos de unidades lexicales clasificadas según criterios lingüísticos o a través de diccionarios y plantillas semánticas declarado por el usuario (Lancia, 2012).

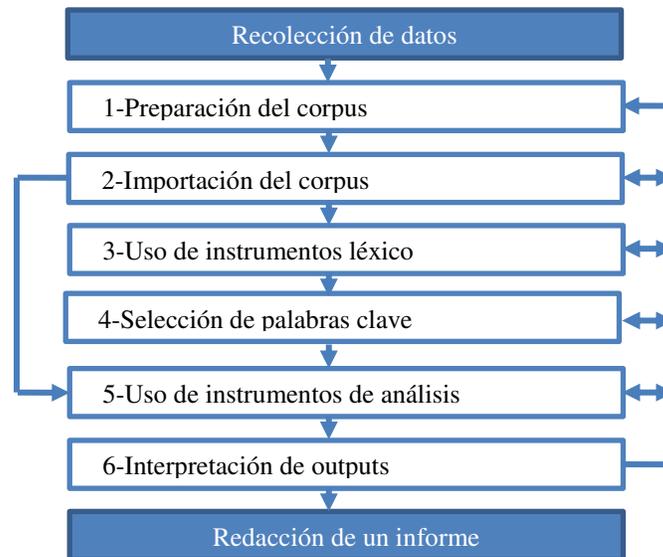
Las “unidades de contexto” son porciones de texto en las que se puede dividir el corpus. En particular, en la lógica T-LAB, las unidades de contexto pueden ser de tres tipos:

1. Documentos primarios correspondientes a la subdivisión "natural" del corpus, o sea a los contextos iniciales definidos por el usuario.
2. Contextos elementales, correspondientes a las unidades sintagmáticas en las cuales cada documento primario puede ser subdividido.
3. Subconjuntos del corpus que corresponden a grupos de documentos primarios atribuibles a la misma categoría o a clústeres temáticos conseguidos a través de específicos instrumentos T-LAB (Lancia, 2012).

T-LAB permite explorar y analizar las relaciones entre las unidades de análisis de todo el corpus o de sus subconjuntos (Stefanello et al., 2010). T-LAB está constituido por operaciones que tienen por objeto el mismo corpus y está organizado por una estrategia y por un plan del usuario. Por lo tanto, inicia con la

recolección de textos a analizar y termina con la redacción de un informe (Cortini & Tria, 2014). La sucesión de las distintas fases está ilustrada en el Diagrama 1:

**Diagrama 1.** Fases de funcionamiento de T-LAB.



Fuente: Manual de Usuario de Lancia. (2017).

T-LAB tiene una ventaja para nosotros los hispanoparlantes, ya que dispone de diccionarios de análisis para las siguientes lenguas: inglés, italiano, alemán, portugués, francés, español y latín (Cortini & Tria, 2014).

### **ALCESTE**

De acuerdo a sus creadores, ALCESTE (*Analyse des Lexèmes Co-occurrents dans les Enoncés Simplifiés d'un Texte*) es un software para el **Análisis de Datos Lexicales**, del CNRS, con el apoyo de la ANVAR. Es la propiedad de la Sociedad IMAGEN. La Sociedad IMAGEN está especializada en matemáticas aplicadas y en el desarrollo de software científico y de apoyo a la toma de decisiones (La société IMAGE, 2015).

ALCESTE es un método informatizado para el análisis de textos, creado por Max Reinert (1986) basada en el desarrollo de métodos de análisis de datos lingüísticos, iniciada por Benzécri (1981) en Francia (de Alba, 2004).

ALCESTE es un software para el análisis lexical que ofrece a los investigadores en ciencias sociales el examen de un corpus de texto escaneado. (Kalampalikis et al., 2003).

El fin del método propuesto es el de poner en evidencia, a través del análisis de un conjunto de textos, los universos lexicales evocados por los usuarios. La estadística encuentra todo su interés en ello: es la redundancia de las huellas lexicales lo que permite localizar los universos más frecuentes. Un universo

lexical es entonces la huella estadística de un lugar en el vocabulario, lugar frecuentemente habitado por el usuario (Reinert, 2003).

Según Max Reinert (1999), el autor de el método, ALCESTE nos permite asignar los principales lugares en el habla común, el "común" en el sentido de que son más a menudo "habitadas" por los usuarios. "Estos mundos de la palabra, por lo tanto, las huellas puramente semiótico inscritas en la misma materialidad del texto. En sí mismos, son independientes de cualquier interpretación. Pero, ellos no tienen sentido, para un lector, ya sea a través de una actividad de interpretación de una función especial de su propia lectura" (Reinert, 2003).

**Figura 2.** Interfaces del software ALCESTE.



Las principales herramientas de ALCESTE (Reinert, 1986):

- Análisis del vocabulario, la categorización, la gramática, la reducción de palabras y edición de diccionarios.
- Análisis multilingüe (francés, inglés, portugués, italiano, alemán y español).
- Clasificación de *top-down*, *bottom-up*, minería de texto, clasificación temática, manual de indización de texto.
- Coloración y priorización de las clases por orden de importancia, el análisis temporal.
- Acceso permanente a los textos originales en interacción con su diccionario.
- Análisis del Factor de coordenadas, las correlaciones, las contribuciones.
- Vocabulario de las variables de la cruz-las correlaciones de las formas.
- Curvas para incrementar el vocabulario por categoría gramatical, por individuo y por clase.
- Análisis Temporal, los discos de cerca disco de concordancias.
- Cruce de variables, o de palabras en varios niveles, simplificado o completa.
- Asignación del texto por parte de la clase, por los individuos, y por unidades de texto.
- Concordancia de las formas y segmentos repetidos.

- Comparación de los dos textos, dos de tratamiento, dos discursos.
- Informe detallado en el resumen del informe, el micrófono, la comparación de los individuos, las clases y los términos y condiciones de las variables.
- Sticky, ALCESTEGraph, ALCESTEWriter.
- Configuración simplificada y experto.
- Sinónimos y palabras para excluir.
- Módulo de formato y de limpieza del corpus, de la biblioteca de variables.
- Nubes de palabras, secuencias de palabras y redes factor de clase, individuales y modalidad.
- Redes de formularios de clase, variable, así como para el cuerpo (La société IMAGE, 2015).

El análisis principal se realizará a partir del conjunto de “palabras principales” (sustantivos, adjetivos, verbos, algunos adverbios), que serán a su vez sometidas a una reducción (a sus raíces), deviniendo así “morfemas lexicales” (Peyrat-Guillard, 2006). El objetivo de esta reducción es el de enriquecer lo más posible las relaciones estadísticas implicadas en la concordancia de las formas simples, pues elimina la variabilidad de formas de una palabra, privilegiando únicamente su significado esencial (Guerrero et al., 2009). La elección de esta reducción se debe al procedimiento estadístico que sirve a la clasificación de unidades de contexto: la clasificación jerárquica descendiente (similar al *clúster analysis* desde el punto de vista conceptual, aunque no matemático), que es una técnica derivada del análisis factorial de correspondencias destinado al tratamiento de tablas binarias casi vacías, con más de 90% de ceros (de Alba, 2004).

ALCESTE trabaja con cuatro fases, subdivididas en operaciones (el plan de análisis) (Delavigne, 2014):

1. En el primer paso, ALCESTE reconoce la u. c. i. (unidades de contexto elegido por el analista) y las palabras "*star*" (las variables). A continuación, realiza tres sucesivos procesos: corte el corpus en formas, se procede a una clasificación de los sintáctica y a una lematización (Reinert, 2011).
2. El segundo paso es cortar el cuerpo como una unidad de contexto elemental (aproximadamente, frases) y la clase de acuerdo a su distribución. ALCESTE es así como clases sobre la base de las léxicas de contenido de cada u. c. e.: trae la u. c. e., que contiene las mismas formas léxicas (De Looze, Roy, Coronini, Reinert, & Jouve, 1999).
3. En el tercer paso, los resultados de los cálculos realizados anteriormente se implementan en forma de archivos: las clases que se obtienen, las formas más frecuentes de cada uno de ellos (Reinert, 2011).
4. En el paso cuatro, los cálculos complementarios se realizan:
  - a. dos tipos de test cruzados (una parte del texto se cruza con una variable o una palabra en particular);

- b. un análisis factorial de correspondencias (cruce de vocabulario y de las clases), que puede verse con la ayuda de las representaciones gráficas son muy revelador;
- c. ascendente de clasificación jerárquica que muestra las relaciones más o menos de cerca que las palabras tienen el uno con el otro (Reinert, 2003)

Estos análisis son una ayuda para la interpretación de los resultados de la estadística y la descripción de las clases (Kalampalikis et al., 2003).

ALCESTE fue diseñado originalmente para tratar textos en lengua francesa, ha sido adaptado para utilizarse en varios idiomas: español, inglés, italiano y portugués (La société IMAGE, 2015).

### IRAMUTEQ

IRAMUTEQ (*Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires*) es un software libre (GNU), creado con programación Phyto para utilizarse con el software R de análisis estadístico. Esto quiere decir que para poder utilizar el software tendremos que instalar primero este otro (Souza, Wall, Peres, Thuler, & Lowen, 2018).

Este software fue desarrollado por el francés Pierre Ratinaud en 200, un investigador de la Universidad de Toulouse, que trabaja en el laboratorio LERASS (*Laboratoire d'Études et de Recherches Appliquées en Sciences Sociales*), en Francia (Ratinaud, 2019).

Al igual que el resto de los softwares anteriores fue desarrollado para realizar análisis de textos en forma multidimensional bajo la utilización corpus y análisis de correspondencias (Ruiz Bueno, 2017).

**Figura 3.** Interface inicial del Software IraMuTeQ



El software IRAMUTEQ utiliza el mismo algoritmo creado por Reinert para su software ALCESTE, para realizar el análisis del corpus textual (Ratinaud, 2019).

Las etapas o herramientas utilizadas por IRAMUTEQ son: (Ruiz Bueno, 2017)

- **Reducción de las unidades del corpus textual.** UCE (Unidades de Contexto Elementales) representan un texto de dos o tres líneas. El objetivo de este análisis es clasificar estas UCE.
- **Lematización.** Los textos (palabras) son lematizados, es decir, reducidos a su raíz (Lemaire, 2008).

- **Creación de unidades de contexto (UC).** La unidad de contexto es un conjunto de UCE que cuenta con un número mínimo de "formas activas " (verbos, nombres, adverbios, adjetivos) por oposición a "formas suplementarias" o "palabras-herramientas" (preposiciones, pronombres, adjetivos posesivos, ciertos verbos y adverbios frecuentes).
- **Clasificación jerárquica descendente (CHD).** Es un clúster jerárquico descendiente partiendo de un grupo o clase, creado por un análisis factorial de correspondencias y que incluye a todos los individuos.
- **Análisis de grafos.** Es un estudio sobre la proximidad y relación de los elementos del conjunto pero de forma reductiva, hasta quedar sin enlaces.

En IRAMUTEQ la preparación de datos es muy importante, dado que el software es libre, es necesario darle al mismo toda la codificación previa para que el software pueda trabajar con él. Se debe dar formato al texto utilizando además otro software específico: OpenOffice (Software libre también) y codificarlo en versión UTF-8 (8-bit Unicode Transformation Format). De esta forma se insertan las variables y el tipo de texto que estamos ingresando al analizador (Moreno & Ratinaud, 2015):

- a. Texto sin variables categoriales asociadas
- b. Texto con variables categoriales identificadas (un tema)
- c. Texto con varios temas y variables categoriales asociadas

Uno de los principales atractivos realizado por IRAMUTEQ, y que fue propuesto inicialmente por el software ALCESTE es el análisis de Clasificación Jerárquica Descendente DHC (*Descending Hierarchical Classification*) (Souza et al., 2018), que muestra, por medio de un algoritmo divisivo en forma inversa a los ascendentes, la clasificación jerarquizada de los grupos e individuos como definidos como clases lexicales, donde cada una de ellas representa una temática y pueden ser descritos por el vocabulario que los define (Ruiz Bueno, 2017).

El Análisis Factorial de Correspondencia refuerza y complementa estos resultados dado que categoriza la valencia de las emociones en un espectro que va de las positivas a las negativas, así como categoriza el tipo de tareas según si son individuales o colectivas (Moreno & Ratinaud, 2015).

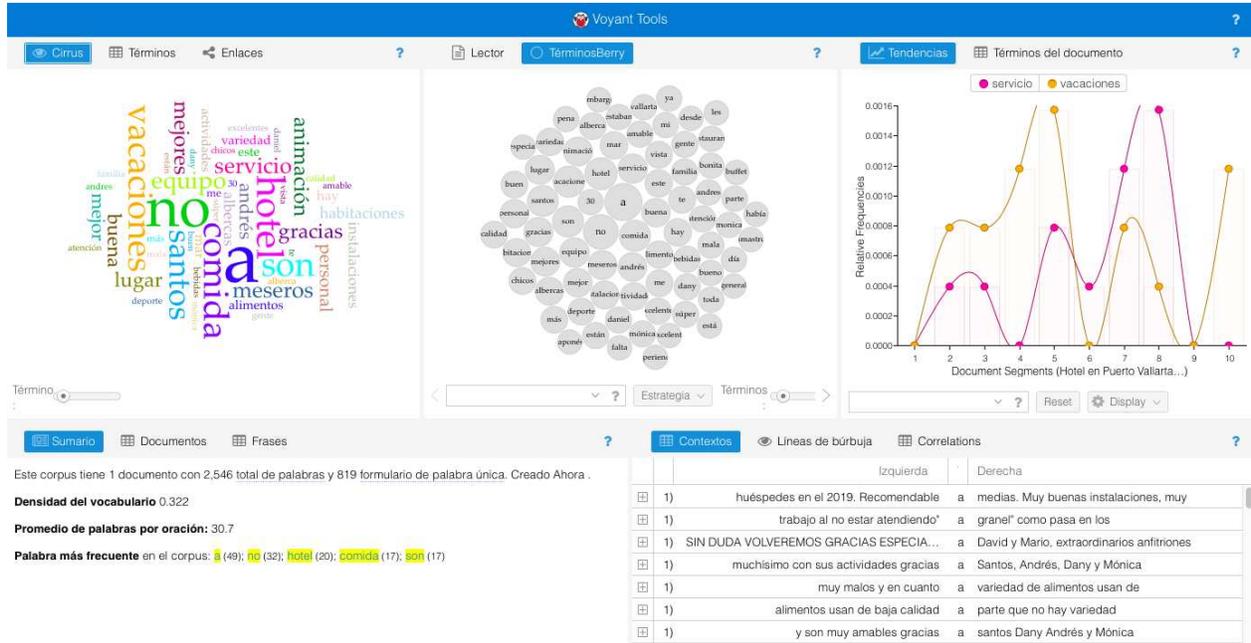
El software IRAMUTEQ está disponible en 4 idiomas: francés, inglés, español y portugués. La última actualización fue realizada el 22 de diciembre del 2014, y va en su versión 0.7 alpha 2.

## **VOYANT-TOOLS**

VOYANT-TOOLS es un software libre (*open source*), anidado en un sitio web, para el análisis de textos y herramientas de visualización creado por Stéfán Sinclair y Geoffrey Rockwell (Sinclair, Ruecker, & Radzikowska, 2013).

VOYANT-TOOLS es una web basada en la lectura del texto y análisis del entorno. Está diseñado para facilitar la lectura e interpretación de las prácticas de las humanidades digitales, estudiantes y académicos, así como para el público en general (Welsh, 2014).

**Figura 4.** Interface del sitio web voyant-tools.org



VOYANT-TOOLS sirve para:

- Utilizarla para el estudio de los textos que se encuentran en la web o textos que se han cuidadosamente editado desde su equipo.
- Se usa para añadir funcionalidad a las colecciones en línea, revistas, blogs o sitios web para que otros puedan ver a través de sus textos con las herramientas de análisis.
- Se usa para añadir interactivo pruebas de sus ensayos que publica en línea. Agregar paneles interactivos a la derecha en su investigación ensayos (si pueden ser publicados en internet) para que sus lectores puedan recapitular los resultados (About VOYANT-TOOLS, 2017).

Aunque con esta herramienta de análisis de los textos puede elegir resaltar los diferentes aspectos de tal propósito, las principales características de VOYANT-TOOLS, a diferencia de otras herramientas (Sinclair & Rockwell, 2015):

- Separación de interfaz: lo mejor puede ser separado del *back-end* de procedimientos analíticos del *front-end* de la interfaz, sin preocupaciones
- Modularidad: las herramientas deben ser capaces de encajar en varias configuraciones
- Inter-operabilidad: las herramientas deben proporcionar las APIs públicas para que puedan interactuar con otras herramientas en la web

- Sencillez: al menos una vista de las herramientas debe ser máximamente simple en su interfaz
- Flexibilidad: las herramientas deben ser capaces de trabajar con archivos locales o de la red, y en diferentes formatos
- Referenciable: la herramienta y sus resultados se prestan a que sea referenciado y citado como recurso académico.
- Internacionalización: las herramientas deben permitir que los usuarios que trabajen en diferentes idiomas
- Rendimiento: las herramientas deben ser razonablemente sensibles para funcionar en un contexto de sitio web
- Extensibilidad: debe ser fácil para crear nuevas herramientas y adaptación de las ya existentes, especialmente para los fines de experimentación propios
- Presentable (*skin*): las herramientas deben ser capaces de presentarse (cambiar apariencia) a sí mismos de forma diferente para diferentes necesidades de los usuarios y preferencias
- Dominio de la sensibilidad: las herramientas deben ser sensibles a las formas en que piensan los eruditos textuales y de interactuar con los textos digitales
- Escalabilidad: herramientas que proporcionan una funcionalidad tanto para un corpus pequeño (como un libro) o un corpus grande (como muchos libros)
- Presencia: herramientas que se prestan para ser incorporados a los contenidos de cualquier otro lugar en la web
- Generalización: las herramientas deben estar diseñadas para abordar una variedad de tipos de texto y usos

VOYANT-TOOLS, como su nombre lo indica, se diferencia en que tiene diferentes herramientas, las cuales se listas a continuación con su uso principal (GitHub, 2017):

1. *Bubblelines*: visualiza la frecuencia y repetición de una palabra es su uso en un corpus. Cada documento en el corpus se representa como una línea horizontal y se divide en segmentos de igual longitud. Cada palabra seleccionada se representa como una burbuja con el tamaño de la burbuja indica la frecuencia de la palabra en el correspondiente segmento de texto. La más grande la burbuja de la radio con más frecuencia la palabra se produce.
2. *Bubbles*: es una visualización de las frecuencias de término en el documento.
3. *Cirrus*: es una nube de palabras que permite visualizar la parte superior de la frecuencia de palabras de un corpus o documento.
4. *Collocates Graph*: representa las palabras clave y los términos que se producen en la proximidad como una fuerza dirigida gráfico de la red.

5. *Corpus Collocates*: es una vista de tabla de los términos que aparecen con más frecuencia en la proximidad de las palabras clave a través de la totalidad del corpus.
6. *Contexts*: (o palabras claves en Contexto), herramienta que muestra cada una de las ocurrencias de una palabra clave con un poco de texto que lo rodea (el contexto).
7. *Correlations*: permite una exploración de la medida en que las frecuencias de términos varían en sincronización (términos cuyas frecuencias suben y bajan juntos, o a la inversa).
8. *Document Terms*: es una vista de tabla de documento de las frecuencias de término.
9. *Corpus Terms*: es una vista de tabla de las frecuencias de término en todo el corpus.
10. *Documents*: se muestra una tabla de los documentos en el corpus, e incluye una funcionalidad para modificar el corpus.
11. *Knots*: es una visualización creativa que representa en términos de un único documento como una serie de líneas torcidas.
  1. *Mandala*: es una visualización conceptual que muestra las relaciones entre los términos y documentos.
  2. *Microsearch*: visualiza la frecuencia y la distribución de los términos en un corpus.
  3. *Phraes*: muestra las secuencias de repetición de palabras organizado por la frecuencia de repetición o el número de palabras en cada frase repetida.
  4. *Reader*: proporciona una forma de lectura de documentos en el corpus, el texto se tomó de sobre-demanda según sea necesario.
  5. *Scatterplot*: es una gráfica de visualización de cómo las palabras se agrupan en un corpus uso de similitud de documentos, el análisis de correspondencias o análisis de componentes principales.
  6. *StreamGraph*: es una visualización que muestra el cambio de la frecuencia de las palabras en un corpus (o dentro de un solo documento).
  7. *Summary*: ofrece una simple, textual visión general de la actual corpus, incluyendo la información acerca de las palabras y de los documentos.
  8. *TermsRadio*: es una visualización que muestra el cambio de la frecuencia de las palabras en un corpus (o dentro de un solo documento).
  9. *TextualArc*: es una visualización de los términos de un documento que incluye un centroide ponderado de términos y un arco que sigue los términos en el orden del documento.
  10. *Topics*: proporciona una forma rudimentaria de la generación de plazo clústeres de un documento o de corpus y, a continuación, ver cómo cada tema (término clúster) se distribuye en todo el documento o corpus.
  11. *Trends*: las tendencias se muestra un gráfico que representa la distribución de una palabra de la ocurrencia a través de un corpus o documento.

12. *Veliza*: es una herramienta experimental para tener una (limitada) natural de intercambio de idiomas (en inglés) basado en el corpus.

13. *Word Tree*: es una herramienta que permite explorar cómo se utilizan las palabras en frases.

Las diferentes herramientas en la interfaz están diseñadas para interactuar el uno con el otro. La interactividad y la navegación entre las diferentes escalas de un corpus son una parte clave del diseño de VOYANT-TOOLS (Sinclair & Rockwell, 2015).

La interfaz de usuario de VOYANT-TOOLS está disponible en varios idiomas. Por defecto VOYANT-TOOLS detecta el idioma del navegador y presenta la interfaz en el idioma de preferencia (o en inglés si no hay otros disponibles). Los siguientes idiomas están disponibles (en varias etapas): árabe, bosnio, croata, checa, inglés, francés, hebreo, italiano, japonés y serbio (VOYANT-TOOLS Help, 2017).

### **Resultados de la evaluación práctica**

Después de un análisis comparativo de los software, sustentando en la información bibliográfica proporcionada por manuales e investigadores usuarios, pasaremos a valorar de manera general los diferentes productos analizados. Es difícil concretar cuál de ellos es el producto más completo, los cuatro tienen características relevantes de las que los otros no y viceversa. Hay que decir que hay aspectos destacados en cada uno de ellos: ALCESTE tiene una particularidad básica que es sencillo de utilizarse al tener una interfaz intuitiva y simplificada, algo que se aprecia para los que empiezan a utilizar el software; T-LAB aventaja en su capacidad para trabajar con modos experto o principiante, lo que tiene un nivel de personalización para trabajar de acuerdo a la curva de aprendizaje del mismo; T-LAB destaca, a nuestro parecer, por su uso de barras 3D dinámicas en html, además de contar con recursos en español como manuales de ayuda e instalación. Hay que señalar que ambos han sacrificado buena parte de las posibilidades de customización que podrían tener los software de versión libre u *open source* que son más difícil de utilizar, pero modulares en su adecuación personalizada (Cortini & Tria, 2014).

VOYANT-TOOLS por su parte es accesible completamente al estar montado en un sitio web y ser en formato libre, para su uso gratuito por cualquiera que tenga un navegador instalado y acceso a internet. Pero, así como es libre, también tiene la limitación de herramientas y cantidad de documentos analizados.

En primer lugar, se empieza con una evaluación general de los cuatro, seguido por los aspectos más específicos. En la Tabla 2, se muestran las generalidades de cada uno de los software, como son la empresa o institución que los creó y/o comercializa actualmente; el año de aparición que indica la antigüedad del mismo, así como un nivel de experiencia como positivo, o una desactualización como negativo; la última actualización o versión que ha sido lanzada al mercado; el costo de los que son de paga, que son los dos de Europa; el tipo de licencia que ofrecen al mercado; el sistema operativo sobre el que se pueden instalar o “correr”; el tipo de archivos desde los cuales se puede importar a este software

para trabajar el corpus; y por último, el tipo de archivo de salida que nos generan para poder trabajarlo en las investigaciones, llamada exportación.

**Tabla 2.** Evaluación de aspectos generales

| Software     | Compañía          | Año de aparición | Última versión | Coste                 | Licencia                | Sistema operativo                 | Importación   | Exportación                         |
|--------------|-------------------|------------------|----------------|-----------------------|-------------------------|-----------------------------------|---|-------------------------------------|
| ALCESTE      | La Sociedad IMAGE | 1986             | 2018           | \$4,406 <sup>2</sup>  | Empresarial y Educación | Windows                           | .txt, .xls, .doc,   | Png,                                |
| IRAMUTEQ     | LERASS            | 2008             | 2014           | Gratis                | GNU software libre      | Windows, Macintosh OS X, y Linux. | utf8, .csv  | .ira .png y .svg                    |
| T-LAB        | T-LAB             | 2001             | 2019           | \$29,230 <sup>3</sup> | Estándar y Educación    | Windows 7 o superior              | .txt, .doc, .docx, .pdf, .rtf, .html, .xls, .xlsx, .csv, .mdb, .accdb | html, emf, wmf, bmp, jpg, png, txt. |
| VOYANT-TOOLS | GitHub            | 2003             | 2.2            | Libre en sitio web    | <i>Creative Commons</i> | Ambas Sobre navegador             | .html, .xml, .pdf, .rtf, MS Word, Excell, ODT y pages (Apple).        | Html, xls y png                     |

Fuente: Elaboración propia.

La Tabla 3 no trata las generalidades, es una evaluación de los aspectos más específicos y que hacen de cada software especial y único en su tipo, que los hace expertos en el análisis de semántica lexical. En primer lugar se comparan los diversos tipos de idiomas en los que se puede trabajar, ya que el principal tema es el análisis de textos en su idioma original, este es uno de los principales plus; se añade una columna de limitaciones que se detectó, de acuerdo al autor, al usar los cuatro para el análisis de corpus; otra característica es la compatibilidad entre otros software utilizados en el ámbito de los investigadores, ya que por lo general vamos a utilizar estos datos para darle otra salida o uso; y por supuesto, no podía falta una lista de las principales herramientas ofrecidas por cada uno de ellos; otro aspecto es el tipo de gráficas generadas por cada uno, debido a que los resultados se deben graficar para su presentación final en artículos científicos, es uno de los aspectos importantes a tomar en cuenta; y finalmente, el tipo de análisis “sugerido” para el que fueron hechos de acuerdo a sus creadores.

<sup>2</sup> El costo es de € 199, que a la fecha de realizar el artículo se ha realizado la conversión a pesos mexicanos. Se toma el precio de € 1 = 22.14414 MXN

<sup>3</sup> El costo es de € 1,320, que a la fecha de realizar el artículo se ha realizado la conversión a pesos mexicanos. Se toma el precio de € 1 = 22.14414 MXN

**Tabla 3.** Evaluación de aspectos específicos

| Software            | Lenguas (idiomas)   | Limitaciones principales                          | Compatibilidad                          | Herramientas  | Gráficas   | Análisis de:  |
|---------------------|---|---|---|---|--|---|
| <b>ALCESTE</b>      | español, inglés, italiano y portugués   | Tamaño de texto a ingresar mínimo                 | SPSS y R                                | -Análisis de vocabulario<br>-Análisis factorial<br>-Análisis temporal<br>-Cartografía de textos<br>-Concordancia<br>-Comparación de dos textos.<br>-Análisis de correlaciones   | -Colorización y clasificación<br>-Cantarell<br>-Red<br>-Disco de proximidad<br>-Disco de concordancias<br>-Barras                | -Textos<br>-Informes de reuniones<br>-Obras literarias<br>-Preguntas abiertas de entrevistas  |
| <b>IRAMUTEQ</b>     | Francés, portugués, inglés, italiano y español                                      | No hay soporte.                                   | R, Python y Alceste<br>Tropes y OwlEdge | -Clusters<br>-Análisis de frecuencias<br>-Análisis de especificidades<br>-Clasificación jerárquica descendente (método Reinert)<br>-Análisis de similitud<br>-Nubes de palabras<br>-Análisis Factorial de Correspondencia | -Gráficos 2D y 3D<br>-Corpus<br>-Nubes de palabras<br>-Dendogramas<br>-Matriz de similitud<br>-Frecuencias<br>-Nubes de palabras | -Textos<br>-Entrevistas<br>-Cuestionarios<br>-Páginas web<br>-Noticias<br>-Leyes<br>-Preguntas abiertas   |
| <b>T-LAB</b>        | inglés, italiano, alemán, portugués, francés, español y latín.                      | 50 categorías por variable                        | Excel, Word                             | -Análisis de temática<br>-Análisis de Co-correncia<br>-Análisis comparativo   | -Barra<br>-Pie<br>-Diagramas<br>-Barras 3D<br>-Clústeres<br>-Concordancias-<br>Gráficos dinámicos html<br>-Tree map              | -Artículos de periódicos<br>-Transcripciones de entrevistas y discursos<br>-Respuestas a preguntas abiertas<br>-Mensajes de twitter<br>-documentos empresariales<br>-Textos legislativos<br>-Libros |
| <b>VOYANT-TOOLS</b> | Inglés, arábico, bosnio, checo, croata, francés, hebreo, italiano, japonés y serbio | En línea solamente, No se puede guardar la sesión | Excel                                   | -Visualización de frecuencia y distribución del corpus<br>-Cirrus de la frecuencia de palabras<br>-Tabla de corpus<br>-Correlaciones<br>-Términos del corpus<br>-Visualización de frecuencias por documento               | -Bubblelines<br>-Cirrus<br>-Wordtree<br>-Corpus  | -Literatura<br>-Datos de pacientes<br>-Textos de arquitectura<br>-Documentos de investigación   |

Fuente: Elaboración propia.

### Resultados del panel de expertos.

Como complemento a la triangulación de datos, el tercer complemento es la realización de un panel de expertos, a los cuales se les pidió contestar un cuestionario (Anexo 1), el que principalmente se les pidió calificar su apreciación y experiencia en el uso de los software y el rendimiento de diversos aspectos en escala de Likert y un análisis FODA:

- 1) Funcionalidad
  - a) Rendimiento
  - b) Usabilidad
  - c) Accesibilidad
  - d) Compatibilidad
- 2) Aplicación en análisis lexical
  - a) Gráficas
  - b) Lenguajes (idiomas)
  - c) Herramientas
- 3) Consideraciones generales
  - a) Precio
  - b) Soporte
  - c) Información y manuales

Como resultado de este cuestionario se ha elaborado la Tabla 4, en la que se presentan los puntos fuertes, y en contraste las oportunidades o aspectos mejorables de cada software, como resultado del cuestionario a expertos.

**Tabla 4.** Resultados del FODA de expertos

| <i>Software</i> | <b>Puntos fuertes</b>  | <b>Aspectos mejorables</b>  |
|-----------------|--|---|
| <b>ALCESTE</b>  | <ul style="list-style-type: none"> <li>-Versión en computadora local</li> <li>-Interfaz colorida y moderna</li> <li>-Uso de método Reinert de análisis lexical</li> <li>-Usabilidad</li> <li>-Análisis de corpus individual</li> <li>-Dos formas de reporte: resumido y detallado</li> <li>-Análisis de correlaciones factoriales</li> <li>-Cursos de capacitación presenciales y online</li> <li>-Análisis bastante completo y robusto</li> <li>-Adecuado para investigación académica</li> <li>-Uso recurrente por investigadores profesionales</li> </ul> | <ul style="list-style-type: none"> <li>-Limitaciones en el número de texto</li> <li>-No hay versión en línea</li> <li>-Manuales y ayuda en francés</li> <li>-Primero hay que preparar el corpus textual</li> <li>-Interfaces y programa en francés</li> <li>-Personalización de versión</li> <li>-Precio alto</li> <li>-Compatibilidad entre las versiones de Windows</li> <li>-Complejo de entender</li> </ul> |
| <b>IRAMUTEQ</b> | <ul style="list-style-type: none"> <li>-Es completamente gratuito</li> <li>-Uso de método Reinert de análisis lexical</li> <li>-Compatibilidad con todos los sistemas operativos existentes.</li> <li>-Gráficas 2D y 3D</li> <li>-Medición de emociones al igual que ALCESTE</li> <li>-Manual en español</li> <li>-Software de interfaces en 4 idiomas diferentes: francés, inglés, español y portugués.</li> </ul>  | <ul style="list-style-type: none"> <li>-Arreglo de textos para usar con el software a analizar es complicado</li> <li>-Necesario instalar software R para utilizarlo</li> <li>-Uso en más idiomas, no solamente 4</li> <li>-Sin soporte ante dudas o mal funcionamiento</li> <li>-Salidas de archivos y compatibilidad limitada</li> <li>-No hay suficiente actualización</li> </ul>                            |
| <b>T-LAB</b>    | <ul style="list-style-type: none"> <li>-Modo experto y principiante</li> <li>-Gráficas en 3D</li> <li>-90MB de corpus (55,000 páginas)</li> <li>-Trabaja con idiomas cirílicos: chino, ruso y árabe.</li> </ul>  | <ul style="list-style-type: none"> <li>-Máximo de 50 categorías por variable</li> <li>-Incompatibilidad con otros software estadísticos.</li> <li>-No hay versión en línea</li> <li>-Precio alto</li> </ul>   |

| <i>Software</i>          | <b>Puntos fuertes</b>  | <b>Aspectos mejorables</b>  |
|--------------------------|--|---|
|                          | <ul style="list-style-type: none"> <li>-Configuración personalizada</li> <li>-Gráficos dinámicos html</li> <li>-Manuales y ayuda en español.</li> <li>-Análisis completo y estable</li> <li>-Uso sencillo</li> </ul>   | <ul style="list-style-type: none"> <li>-Interfaces y programa en inglés</li> <li>-Interface gráfica no moderna</li> </ul>   |
| <b>VOYANT-<br/>TOOLS</b> | <ul style="list-style-type: none"> <li>-Gratuito</li> <li>-En la nube todo el tiempo</li> <li>-Complementa otros análisis</li> <li>-Interface amigable</li> <li>-Visualización dinámica e interactiva</li> <li>-Usabilidad amplia</li> <li>-Gráficas interactivas</li> <li>-Liga a las gráficas para su visualización posterior</li> <li>-Interface gráfica atractiva</li> <li>-Actualización constante por parte de la comunidad</li> </ul> | <ul style="list-style-type: none"> <li>- Necesidad de conexión a internet</li> <li>- Herramientas de tipo general y no especializado</li> <li>- No se guarda la sesión para posteriores análisis</li> <li>- Dar crédito a los creadores por su uso</li> <li>- Limitación por número de documentos a subir a la nube</li> <li>- Límite en el tamaño de los archivos a subir a la nube</li> <li>- Tiempos largos al subir archivos grandes</li> </ul> |

Fuente: Elaboración propia

Finalmente, pasemos a las conclusiones y recomendaciones a los investigadores interesados en utilizar este tipo de software en sus futuras investigaciones.

### **Conclusiones y recomendaciones**

Los estudios cualitativos priorizan las producciones discursivas, traducidas en una colección de los textos que requieren una preparación y una exposición razonada para su selección, de acuerdo con argumentos epistemológicos, por tanto desde el muestreo y los criterios de inclusión del sujeto, hasta la necesidad de clarificaciones, además de la aseguración de una base lógica, válida, confiable y de confianza para el estudio (Azevedo et al., 2013). Lo que implica tener herramientas confiables a la hora de realizar la elección de un software de este tipo, principalmente si requiere una inversión significativa por parte del investigador o institución en la que labora.

A diferencia de los paquetes tradicionales de análisis cualitativo, el software de análisis lexical ofrecen solamente un medio que permite ajustar el tiempo invertido en función de los objetivos de la investigación, los análisis pueden realizarse fácilmente entre las variables subyacentes en el estudio, y de esta forma, tener un nuevo tipo de información a examinar: los aspectos estructurales del lenguaje (Bolden & Moscarola, 2000).

Es importante considerar que estos software deben de ser meramente herramientas para procesar datos, como lo es el SPSS o STATA, no como un método de investigación, sino como un instrumento de exploración, búsqueda y asociación de material para investigar.

Ahora las conclusiones por cada uno de los software, en primer lugar, ALCESTE es un programa complejo, que aunque versátil, tiene una curva de aprendizaje no tan amigable, sus procesos estadísticos son poderosos: la clasificación jerárquica y el análisis de correspondencia factorial son sus principales herramientas poderosas. Sin embargo, el programa presenta ciertas limitaciones, una de ellas es que al no tomar en cuenta la construcción sintáctica de las frases, sino sólo las palabras reducidas a sus raíces más

frecuentemente asociadas entre sí, se pierde el sentido original del texto y se corre el riesgo de hacer algunas interpretaciones equivocadas (de Alba, 2004). Su parte más fortalecida es que ha sido generado por uno de los autores más reconocidos en el análisis lexical: Max Reinert, y por tanto utiliza el algoritmo generado por él (Reinert, 2011). Su principal fortaleza, por otro lado, es que es el único que cuenta con el respaldo de una institución que capacita indefinidamente en el uso de su software, el inconveniente es que es solamente en París y en idioma francés.

Por otra parte, T-LAB es un software que tiene un precio alto para los estándares de un investigador de universidad pública, pero hay que decir que tiene las características exclusivas que hace que lo valga: sus gráficas en 3D, su configuración personalizada y el dinamismo de sus gráficos html. Pero para su elevado precio, se espera una interface más moderna, algo que parece necesario, pero que no se ha tomado en cuenta. Algo más a su favor, es que su uso es sencillo y tiene una plataforma estable y robusta. Su principal fortaleza es que, tomando las características de los otros softwares en el mercado, ha añadido sus principales herramientas y las ha magnificado y mejorado, eso lo hace uno de los software más completos y complejos ante su competencia. Por ejemplo, en el caso de los análisis de correspondencias, el resto de los software (ALCESTE e IRAMUTEQ) solamente tienen un tipo, mientras que T-LAB tiene tres tipos diferentes (Lancia, 2012). Si el investigador no tiene límite para adquirir recurso de este tipo, es una opción para tomarse en cuenta.

En el caso de IRAMUTEQ, el software realizado por los franceses es cuenta con las funciones básicas y más importantes de ALCESTE, y el algoritmo de Max Reinert. Pero al ser un software de libre licencia (GNU), no cuenta con los recursos para tener su robustez ni las suficientes actualizaciones en forma constante. El principal problema con IRAMUTEQ es su dependencia de otra plataforma para su funcionamiento, en este caso el software R que, aunque goza de fama en la comunidad de investigadores y es una plataforma estable hasta el momento, lo hace requerir recursos externos para funcionar. Su principal fortaleza, por tanto es su gratuidad y robustez en los principales análisis utilizados con el algoritmo Reinert.

Finalmente, VOYANT-TOOLS, no puede competir con las capacidades de los anteriores por su calidad de software *open source*, pero al ser precisamente gratis, tiene la capacidad cambiante y renovable que los caracteriza. Su interfaz es intuitiva y amigable, lo que lo hace de un uso fácil, acompañado de gráficas interactivas. Nuestro consejo para la utilización de este software, es que sea un acompañamiento y complemento a cualquiera de los anteriores en el análisis de textos, porque por sí solo carece de características y herramientas que no podrían generar el análisis completo lexical que ofrecen los de paga. Tal vez en un futuro, si la herramienta se sigue desarrollando, pueda ser una opción completa. Pero por el momento, es una opción nueva, refrescante y sencilla de apoyo.

Al describir detalladamente la utilización de los cuatro software, el presente estudio tenía por objeto destacar algunos aspectos pertinentes y contribuir a la difusión de estos instrumentos, que requieren conocimientos y aptitudes, además de no descuidar la importancia del análisis del investigador y los aspectos éticos inherentes a la investigación cualitativa. En las ciencias de la administración los análisis cualitativos son cada vez más recurrentes y solicitados, lo que hace necesario este tipo de revisiones para apoyo de los investigadores interesados en utilizar estas herramientas.

## Referencias

- About VOYANT-TOOLS | VOYANT-TOOLS Documentation. (2017). Docs.VOYANT-TOOLS.org. Retrieved 22 August 2017, from <http://docs.VOYANT-TOOLS.org/about/>
- Álvarez Cadavid, G. M., Giraldo Ramírez, M. E., & Navarro Plazas, C. del P. (2017). Uso de TIC en investigación cualitativa : *Revista Katharsis*, 23(enero-junio), 218–235.
- AntCon. (2017). Laurence Anthony's Website. Retrieved 14 Dec 2017, from <http://www.laurenceanthony.net/software/antcon/>
- Azevedo, D. M. de, Costa, R. K. de S., & Miranda, F. A. N. de. (2013). Uso do Alceste na análise de dados qualitativos: contribuições na pesquisa em enfermagem. *Revista de Enfermagem UFPE on Line*, 7(esp), 5015–5022. <https://doi.org/10.5205/reuol.4700-39563-1-ED.0707esp201326>
- Baker, P., Hardie, A., & McEnery, T. (2006). *a Glossary of Corpus Linguistics*. Edinburgh: Edinburgh University Press. Retrieved from [https://s3.amazonaws.com/academia.edu.documents/41608722/A\\_Glossary\\_of\\_Corpus\\_Linguistics\\_Glossaries\\_in\\_Linguistics.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1511254931&Signature=9od5PS%2F%2FZc4N%2B86P3%2FQQ%2FBk9i9A%3D&response-content-disposition=](https://s3.amazonaws.com/academia.edu.documents/41608722/A_Glossary_of_Corpus_Linguistics_Glossaries_in_Linguistics.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1511254931&Signature=9od5PS%2F%2FZc4N%2B86P3%2FQQ%2FBk9i9A%3D&response-content-disposition=)
- Bolden, R., & Moscarola, J. (2000). Bridging the quantitative-qualitative divide - The lexical approach to textual data analysis. *Social Science Computer Review*, 18(4), 450–460. <https://doi.org/10.1177/089443930001800408>.
- CasualConc. (2017). *ConsualConc*. Retrieved 14 Dec 2017, from <https://sites.google.com/site/casualconc/Home>
- Colina, L., Pericás, V., Miquel, J., Olivé, M., & Roldán, L. (2003). Relaciones, redes y discurso: revisión y propuestas en torno al análisis reticular de datos textuales. *Revista Española de Investigaciones Sociológicas*, 101, 175–200.
- Cortini, M., & Tria, S. (2014). Triangulating Qualitative and Quantitative Approaches for the Analysis of Textual Materials: An Introduction to T-Lab. *Social Science Computer Review*, 32(4), 561–568. <https://doi.org/10.1177/0894439313510108>
- Crook, T. R., Todd, S. Y., Combs, J. G., Woehr, D. J., & Ketchen Jr, D. J. (2011). Does human capital matter? A meta-analysis of the relationship between human capital and firm performance. *Journal of Applied Psychology*, 96(3), 443. <https://doi.org/10.1037/a0022147>
- de Alba, M. (2004). El método ALCESTE y su aplicación al estudio de las representaciones sociales del espacio urbano: el caso de la Ciudad de México. *Papers on Social Representations*, 13, 1.1-1.20. <https://doi.org/10.1016/j.appet.2008.11.008>
- De Looze, M.-A., Roy, A., Coronini, R., Reinert, M., & Jouve, O. (1999). Two measures for identifying the perception of risk associated with the introduction of transgenic plants. *Scientometrics*, 44(3), 401–426. <https://doi.org/10.1007/BF02458487>
- Delavigne, V. (2014). Alceste , un logiciel d ' analyse textuelle To cite this version : *Textes et Cultures*. Retrieved from <https://hal.archives-ouvertes.fr/hal-00924168>
- Gambetti, R. C., & Graffigna, G. (2010). The concept of engagement. *International Journal of Market Research*, 52(6), 801–826. <https://doi.org/10.2501/S1470785310201661>
- GitHub. (2017). About - Voyant Tools Help. Retrieved June 6, 2017, from <http://voyant-tools.org/>
- Guerrero, L., Guàrdia, M. D., Xicola, J., Verbeke, W., Vanhonacker, F., Zakowska-Biemans, S., ...

- Hersleth, M. (2009). Consumer-driven definition of traditional food products and innovation in traditional foods. A qualitative cross-cultural study. *Appetite*.  
<https://doi.org/10.1016/j.appet.2008.11.008>
- Hernández, R., Fernández, C., & Baptista, M. (2010). *Metodología De La Investigacion*. México, D.F.: McGraw Hill.
- Huisman, M., & van Duijn, M. A. J. (2004). Software for Social Network Analysis. *Models and Methods in Social Network Analysis*, 1–54. <https://doi.org/10.1017/CBO9780511811395.013>
- Iramuteq. (2019). IraMuTeQ WebSite . Retrieved 1 Ago 2019, from <http://www.iramuteq.org>
- Kalampalikis, N., Abric, J., & Kalampalikis, N. (2003). L ’apport de la méthode Alceste dans l ’analyse des représentations sociales.
- Kalampalikis, N., & Moscovici, S. (2005). Une approche pragmatique de l’analyse Alceste. *Les Cahiers Internationaux de Psychologie Sociale*. <https://doi.org/10.3917/cips.066.0015>
- Kuz, A., Falco, M., Nahuel, L., & Giandini, R. (2015). Análisis de Redes Sociales a través de Gephi y NodeXL, 200–214.
- La société IMAGE. (2015). *Alceste 2015. Un logiciel d’Analyse de Données Textuelles*. Toulouse, France. Retrieved from <http://www.image-zafar.com>
- Lancia, F. (2012). The logic of the T-Lab tools explained, (October), 1–29. Retrieved from <http://tlab.it/en/toolsexplained.php>
- Lancia, F. (2017). *T-LAB Plus 2017. Manual del usuario*. Retrieved from <http://www.tlab.it/>
- Lancia, F. (2017). *T-LAB Instrumentos para el Análisis de Textos*. *Tlab.it*. Retrieved 24 May 2017, from <http://tlab.it/es/presentation.php>
- Larruzea-Urkixo, N., Cardeñoso Ramírez, O., & Idoiaga Mondragon, N. (2019). EL ALUMNADO DEL GRADO DE EDUCACIÓN ANTE LAS TAREAS UNIVERSITARIAS: EMOCIÓN Y COGNICIÓN. *Educación XXI*.
- Leonidas Aguirre, J. (2011). Introducción al Análisis de Redes Sociales. *Documentos de Trabajo CIEPP*, 59.
- Logiciel Alceste. (2017). *Image-zafar.com*. Retrieved 24 May 2017, from <http://www.image-zafar.com/Logiciel.html>
- Moreno, M., & Ratinaud, P. (2015). Manual de uso de IraMuTeQ.
- Open Office Web Site. (2019) Descarga del Software Open Office. Retrieved 1 Ago 2019, from <https://www.openoffice.org/es/descargar/>
- Peyrat-Guillard, D. (2006). Alceste et WordMapper : l ’apport complémentaire de deux logiciels pour analyser un même corpus d ’entretiens. *JADT 2006 : 8es Journées Internationales d’Analyse Statistique Des Données Textuelles*.
- RAE. (2017). Real Academia Española. Retrieved May 12, 2017, from <http://www.rae.es/>
- Ratinaud, P. (2019). Iramuteq. Retrieved August 9, 2019, from <http://www.iramuteq.org/>
- Reinert, M. (1986). Les cahiers de l’analyse des données. *NUMDAM*, 4(11), 471–481.
- Reinert, M. (2003). Le rôle de la répétition dans la représentation du sens et son approche statistique par la méthode “ ALCESTE ” 1. *Semiotica*, 147(1/4), 389–420. <https://doi.org/10.1515/semi.2003.100>
- Reinert, M. (2011). Quelques interrogations à propos de l ’ “objet” d’une analyse de discours de type statistique et de la réponse “Alceste.” *Langage et Société*, 90(1), 57–70.  
<https://doi.org/10.3406/lsoc.1999.2897>
- Ruiz Bueno, A. (2017). Trabajar con Iramuteq: pautas.
- SCP, Simple Concordance Program. (2017). *TextWorld.com*. Retrived 14 Dec 2017 from <http://www.textworld.com/scp/index.html>
- Sinclair, S., & Rockwell, G. (2015). Text Analysis and Visualization. *A New Companion to Digital Humanities*, 274–290. <https://doi.org/10.1002/9781118680605.ch19>
- Sinclair, S., Ruecker, S., & Radzikowska, M. (2013). Information Visualization for Humanities Scholars. *Literary Studies in the Digital Age*, (April 2016). <https://doi.org/10.1632/lstda.2013.6>
- Souza, M. A. R. de, Wall, M. L., Peres, A. M., Thuler, A. C. de M. C., & Lowen, I. M. V. (2018). The use of IRAMUTEQ software for data analysis in qualitative research. *Revista Da Escola de Enfermagem*

- Da U S P*, 52, e03353. <https://doi.org/10.1590/S1980-220X2017015003353>
- Stefanello, G., Francisco, A. De, & Carranza, C. (2010). Aplicación y viabilidad de uso del software de Análisis Cuantitativo de Textos TLAB 7 . 1 en el análisis de las representaciones sociales presentes en la web soyborderline . com Application and utility of the software for Quantitative Content Analysis TLA. *Mediaciones Sociales*, 6, 121–142.
- VOYANT-TOOLS. (2017). *VOYANT-TOOLS*. Retrieved 25 May 2017, from <http://VOYANT-TOOLS.org>
- Welsh, M. E. (2014). Review of Voyant Tools. *Collaborative Librarianship*, 6(2), 96–98. Retrieved from <http://go.galegroup.com/ps/i.do?id=GALE%7CA393876396&sid=googleScholar&v=2.1&it=r&linkaccess=fulltext&issn=19437528&p=AONE&sw=w&authCount=1&u=udeg1&selfRedirect=true>
- WordSmithTools. (2017). *Oxford WordSmith Tools 4.0*. Retrived 14 Dec 2017, from <http://lexically.net/wordsmith/version4/index.htm>