

**UNIVERSIDAD DE SALAMANCA**  
**FACULTAD DE FILOLOGÍA**  
**DEPARTAMENTO DE FILOLOGÍA FRANCESA**



**VNiVERSIDAD**  
**D SALAMANCA**

CAMPUS DE EXCELENCIA INTERNACIONAL

**DOCTORADO EN LENGUAS MODERNAS**

**TESIS DOCTORAL**

**PROPOSITION D'INTÉGRATION DES OUTILS DE TAL POUR UNE  
CLASSE DE FLE, À TRAVERS LE CAS COMPLEXE DU CORPUS  
D'ARCHIVES DE LA PRESSE DE LA PREMIÈRE GUERRE MONDIALE**

**[PROPUESTA DE IMPLEMENTACIÓN DE HERRAMIENTAS DE TAL  
PARA UNA CLASE DE FLE, MEDIANTE EL CASO COMPLEJO DEL  
CORPUS DE ARCHIVOS DE LA PRENSA DE LA PRIMERA GUERRA  
MUNDIAL]**

Autora: ANA PAULA DE OLIVEIRA

Directora de Tesis: MARÍA VICTORIA RODRÍGUEZ NAVARRO

SALAMANCA, 2021



**UNIVERSIDAD DE SALAMANCA**  
**FACULTAD DE FILOLOGÍA**  
**DEPARTAMENTO DE FILOLOGÍA FRANCESA**



**VNiVERSIDAD**  
**D SALAMANCA**

CAMPUS DE EXCELENCIA INTERNACIONAL

**DOCTORADO EN LENGUAS MODERNAS**

**TESIS DOCTORAL**

**PROPOSITION D'INTÉGRATION DES OUTILS DE TAL POUR UNE  
CLASSE DE FLE, À TRAVERS LE CAS COMPLEXE DU CORPUS  
D'ARCHIVES DE LA PRESSE DE LA PREMIÈRE GUERRE MONDIALE**

**[PROPUESTA DE IMPLEMENTACIÓN DE HERRAMIENTAS DE TAL  
PARA UNA CLASE DE FLE, MEDIANTE EL CASO COMPLEJO DEL  
CORPUS DE ARCHIVOS DE LA PRENSA DE LA PRIMERA GUERRA  
MUNDIAL]**

Vº Bº

Directora de la Tesis

Autora

Fdo.:

Fdo.:

Dra. Dña. María Victoria Rodríguez Navarro

Ana Paula De Oliveira

SALAMANCA, 2021



*"En una democracia, todo depende del consentimiento del pueblo.*

*Thomas Jefferson*

*"La lengua hace la nación, la nación hace la lengua".*

*Marcellesi & Garbin*

*"Desde la idea a la palabra; desde la palabra a la idea".*

*Julio Casares*



## **Agradecimientos**

Esta investigación se ha desarrollado, a lo largo de los últimos años, en la Universidad de Salamanca. Por esta razón, mi primer pensamiento va, naturalmente, a todos mis compañeros y profesores universitarios que forman parte de nuestro Departamento de Filología Francesa y que, de cerca o de lejos, han contribuido al progreso de esta tesis. Siempre han sido, para mí, una gran fuente de inspiración y admiración.

Quiero agradecer sinceramente, a mi directora de tesis María Victoria Rodríguez Navarro por haber aceptado realizar, a mi lado, todo este trabajo de investigación y por haberme apoyado durante estos largos años de tesis. Estoy muy agradecida por su sabiduría, su generosidad y su orientación atenta, amable y entusiasta.

También me gustaría agradecer a Ana Teresa González Hernández, Cristina Solé Castells, Fabrice Issac, Juan Manuel Pérez Velasco y Xavier-Laurent Salvador por haber aceptado ser los miembros de mi tribunal de tesis.

Agradezco sinceramente a todos los miembros del Lab'TTN (Théories, Textes, Numérique [Teorías, Textos, Digital]) de la Universidad de París XIII - Sorbona Norte su acogida en el laboratorio, donde he aprendido mucho, especialmente en lo que respecta a las humanidades digitales. Estoy especialmente agradecida a Fabrice Issac por las numerosas oportunidades que me brindó, así como al resto de doctores y doctorandos por su solidaridad y su voluntad de establecer fuertes vínculos entre nuestras respectivas universidades, fomentando así el intercambio de conocimientos.

Estoy especialmente agradecido a Stéphane Lamassé por su hospitalidad en la Universidad de París I - Panthéon Sorbonne y por haber hecho posible mi estancia de doctorado. También le agradezco el tiempo que me dedicó para supervisar mis investigaciones en historia y ciencias del lenguaje.

También quiero agradecer a Frederic Manikaros y a Isabel Ludmila Rivero-Vilá por la redacción de los informes de tesis que me permitieron obtener la

mención internacional. Me gustaría agradecerles por su disponibilidad y su valiosa colaboración.

Por último, me gustaría agradecer a mi familia y amigos su apoyo incondicional y su paciencia durante los largos años de mi tesis. Han trabajado en la sombra, pero han contribuido enormemente al éxito de este proyecto. Que estén aquí agradecidos.



## ÍNDICE DE CONTENIDOS

<b>INTRODUCCIÓN</b> .....	<b>27</b>
<b>I. EL TAL EN EL ÁMBITO DE LA DOCENCIA</b> .....	<b>27</b>
1.1. PROBLEMÁTICA.....	27
1.1.1. Itinerario de investigación inicial.....	27
1.1.2. Aplicación de una metodología .....	30
1.1.3 La sostenibilidad de las lenguas en la era digital .....	33
1.2. DE LAS TIC AL TAL .....	37
1.2.1. Tecnología digital y didáctica .....	37
1.2.2. El TAL: una nueva realidad lingüística y financiera.....	44
1.2.3 Humanidades digitales .....	48
1.2.4. Límites y perspectivas.....	50
1.3. ESTADO DE LA CUESTIÓN.....	51
1.3.1. Panorama de algunos programas aplicables a la enseñanza .....	53
1.3.1.1. Iramuteq .....	53
1.3.1.2. Nooj.....	53
1.3.1.3. Hyperbase.....	54
1.3.1.4. Lexico 5.....	54
1.3.1.6. TXM.....	55
1.4. NUESTRA HERRAMIENTA DE ENSEÑANZA: TXM.....	55
1.4.1. Un analizador de textos en libre acceso y hecho en francia .....	55
1.4.2. Principales funciones de TXM.....	57
1.4.3. Etiquetado morfosintáctico .....	58
1.4.4. Límites de TXM .....	62
<b>II. EL CASO COMPEJO DEL CORPUS DE ARCHIVOS</b> .....	<b>65</b>
2.1. REFERENTE A NUESTRO CORPUS .....	65
2.1.2. Justificación de nuestro corpus .....	65
2.1.3. Constitución de nuestro corpus.....	70
2.1.3. Dificultades relacionadas con el corpus de archivos.....	72
2.1.5. Repertorio de errores resultantes de la OCR.....	77
2.1.6. Importación de nuestro corpus y metodología a seguir.....	82

2.2. INTRODUCCIÓN HISTÓRICA .....	89
2.2.1. Hitos históricos .....	89
2.2.2. Europa en 1914.....	91
2.2.3. Francia en 1914 .....	95
2.2.4. La prensa francesa en 1914.....	99
<b>III. PRINCIPALES OPERACIONES INFORMÁTICAS.....</b>	<b>102</b>
3.1. MARCO METODOLÓGICO Y PRINCIPALES OPERACIONES INFORMÁTICAS .....	102
3.1.1. Principales categorías gramaticales que surgen del corpus .....	102
3.1.2. Los sustantivos .....	107
3.1.3. El predicado .....	112
3.1.4. Análisis del predicado nominal y del verbo de apoyo.....	122
3.1.5 La palabra techo y su marco fraseológico.....	128
3.2. LAS CONSTRUCCIONES FIJAS Y SEMIFIJAS .....	130
3.2.1. Construcciones fijas .....	130
3.2.2. Construcciones semifijas.....	134
3.2.3. Construcciones libres .....	145
3.2.4. La palabra con mayor frecuencia: ¿un vector ideológico ? .....	152
3.3. LAS OTRAS PALABRAS CON MAYOR FRECUENCIA .....	155
3.3.1. Los sustantivos .....	155
3.3.2. Los adjetivos .....	163
3.3.3. Los verbos.....	166
3.4. ANÁLISIS PRAGMÁTICO: LAS INFERENCIAS .....	173
3.4.1. La subjetividad del discurso .....	173
3.4.2. Las analogías.....	178
3.4.3. La sinonimia .....	184
3.4.4. Los antónimos.....	190
3.4.5. La polisemia .....	194
3.4.6. Las relaciones concesivas .....	198
3.4.7. El discurso directo e indirecto .....	202
3.5. ANÁLISIS FACTORIAL DE CORRESPONDENCIAS (AFC) .....	210
3.5.1. Análisis factorial de correspondencias de los sustantivos.....	211
3.5.2. Análisis factorial de correspondencias de los adjetivos .....	215
3.5.3. Análisis factorial de los pronombres personales .....	218

3.5.4. Análisis factorial de los verbos .....	223
3.5.5. AFC de los sustantivos, adjetivos y verbos .....	227
3.6. ESTUDIO SINCRÓNICO.....	230
3.6.1. Análisis sincrónico de los sustantivos "hombres" y "soldados".....	233
3.6.2. Análisis sincrónico del frpos sustantivo y adjetivo .....	236
3.6.3. Análisis sincrónico de los sustantivos "guerra" y "paz".....	238
3.6.4. Análisis sincrónico de los sustantivos en el discurso patriótico .....	240

#### **IV. EXPLOTACIÓN DEL CORPUS MEDIANTE ACTIVIDADES**

##### **SEMIAUTOMATIZADAS EN ASC ..... 245**

4.1. APRENDIZAJE SOBRE CORPUS (ASC).....	245
4.1.2. Marco Común Europeo de Referencia para las Lenguas (MCERL)...	246
4.2. MORFOSINTAXIS .....	249
4.2.1. De la propiedad de los sustantivos .....	249
4.2.2. La formación de sustantivos inanimados: prefijos, sufijos, afijos.....	249
4.2.3. Sustantivos derivados o familias de palabras.....	251
4.2.4. Formación de adjetivos.....	255
4.2.5. Adjetivos a partir de verbos .....	256
4.2.6. Adjetivos con prefijo.....	257
4.2.7. El complemento del sustantivo y la subordinada relativa .....	257
4.2.8. Los adjetivos .....	258
4.2.9. Las construcciones adjetivas .....	259
4.2.10. Ampliación del sustantivo: la cláusula subordinada relativa .....	259
4.2.11. Los artículos.....	262
4.2.12. Los demostrativos y posesivos .....	265
4.2.13. Los indefinidos .....	267
4.2.14. Los pronombres .....	268
4.2.15. El verbo.....	273
4.2.16. Las perífrasis verbales.....	274
4.2.17. Los tiempos y los modos verbales.....	275
4.3. LA FRASE .....	278
4.3.1. Las diferentes formas de negación .....	278
4.3.2. La negación léxica .....	281
4.3.3. La frase interrogativa y la frase exclamativa.....	283

4.3.4. El énfasis y el pleonismo .....	290
4.3.5. Expresión del pensamiento .....	293
4.3.6. Conjunciones coordinadas y subordinadas .....	296
4.3.7. Expresión de las circunstancias .....	297
4.3.8. Expresión del tiempo .....	303
4.3.9. Expresión de oposición y concesión .....	306
4.3.10. Expresión de la hipótesis y la condición .....	308
4.3.11. Expresión de la comparación .....	310
4.3.12. Conectores y marcadores lógico-temporales y del discurso .....	312
4.4. Léxico .....	313
4.4.1. Inventario del léxico .....	314
4.4.2. Campos léxicos .....	316
4.4.3. Ficha léxica .....	319
4.4.4. Aplicación didáctica del léxico .....	324
4.5. LA ORTOGRAFÍA .....	326
4.5.1. Las consonantes dobles .....	327
4.5.2. Los acentos .....	329
4.5.3. Las palabras invariables terminadas en "s" .....	330
4.5.4. La cedilla .....	331
4.5.5. El caso de la letra "g" .....	332
4.5.6. El caso de la letra "h" .....	333
<b>V. CONCLUSIONES (EVALUACIÓN Y PERSPECTIVAS) .....</b>	<b>335</b>
5.1. SÍNTESIS DE LA ARGUMENTACIÓN .....	335
5.2. RESOLUCIÓN DE LA PROBLEMÁTICA .....	336
5.3. POSICIÓN Y CONTRIBUCIÓN .....	341
5.4. LÍMITES .....	341
<b>VI. REFERENCIAS BIBLIOGRÁFICAS .....</b>	<b>343</b>
<b>VII. ANEXOS .....</b>	<b>359</b>
<b>VIII. APÉNDICE PARA LA MENCIÓN INTERNACIONAL .....</b>	<b>519</b>

# INTRODUCCIÓN

## I. EL TAL EN EL ÁMBITO DE LA DOCENCIA

### 1.1 Problemática

#### 1.1.1. Itinerario de investigación inicial

Esta tesis se inscribe en la enseñanza de la lengua francesa en la universidad y, más concretamente, en la enseñanza del Francés como Lengua Extranjera (FLE) a los estudiantes hispanohablantes. Dos idiomas, una raíz común y, sin embargo, de manera subyacente, muchas dificultades porque más allá de la barrera del idioma viene la barrera lingüística inherente.

Este lenguaje, en el sentido en que lo define Martinet, es decir, en su funcionamiento como instrumento estrictamente rudimentario de comunicación: "*En el habla corriente, el "lenguaje" designa propiamente la facultad que tienen los hombres de oírse mutuamente por medio de signos vocales*" (Martinet, 1980, p. 7), pero también en el sentido de que sirve también de "*soporte del pensamiento*", como indica también Martinet. En efecto, cada individuo tiene una relación estrecha y a veces incluso ambigua con cada una de las palabras que componen nuestra lengua.

Debido al hecho de ser profesora de Francés como Lengua Extranjera desde hace varios años en la Universidad de Salamanca en España, pero también antigua alumna del laboratorio LDI (Laboratoire Lexiques, Dictionnaires, Informatique) de la UFR de París XIII Sorbona Norte, y actualmente doctoranda en la Universidad de Salamanca y en el TTN LAB' (Laboratorio de Textos, Teorías, Numéricos, Estudios Interactivos de Lenguas) de París XIII Sorbona Norte, siempre he centrado mis estudios e investigaciones en torno a tres áreas principales: los idiomas, la didáctica y las nuevas tecnologías, porque siempre me ha parecido que estos tres

ámbitos se atraían mutuamente y sólo podían ser significativamente complementarios.

Para mí, las lenguas siempre han sido el caldo de cultivo de muchas reflexiones, ya sea a nivel de la primera lengua o de la traducción. Estas diferentes especialidades siempre han despertado mi curiosidad, sin duda por la complejidad y, por tanto, el reto de interpretar el lenguaje o traducir; Es decir, plasmar una palabra en su significado más cercano o exacto, a pesar de algunos elementos desconocidos o equívocos, porque como muy bien señala Isabel Rivero Vila, *"En un momento en que el planeta se globaliza y los intercambios se multiplican, la posibilidad de malentendidos se incrementa y la necesidad de un aprendizaje de la lengua ligado a la competencia sociocultural es fundamental"* (Rivero Vilá, 2011: 37).

Sabiendo que cada palabra utilizada puede ser en sí misma cuestionada, puesto que ya transmite intrínsecamente una parte íntima del autor, sus concepciones, ideales, principios, pero también su cultura y tradiciones, la ambigüedad aumenta. Como ya dijo Saussure: *"Alguien pronuncia la palabra desnudo: un observador superficial estará tentado de ver en ella un objeto lingüístico concreto; pero un examen más atento llevará a encontrar en ella tres o cuatro cosas perfectamente diferentes, según la forma en que se considere [...] Lejos de que el objeto preceda al punto de vista, parecería que es el punto de vista el que crea el objeto"* (Saussure, 1978, p. 23)

Esto nos lleva evidentemente a la arbitrariedad del signo de Saussure, que subraya la perspectiva del "punto de vista" y, por tanto, el aspecto relativo del significante y del significado *"[...] la elección que uno u otro fragmento acústico requiere para una u otra idea es perfectamente arbitraria. No hay ninguna relación causal lógica o natural que influya en el sistema lingüístico desde el exterior: la unión de la idea y el sonido es radicalmente arbitraria"* (Saussure, 1985, p. 18).

Esta misma reflexión sobre el signo la recoge Carmen García Cela en sus comentarios sobre Saussure: *"Todo el mecanismo del lenguaje depende de él [la linealidad]. A diferencia de los significantes visuales (señales marítimas,*

etc.) que pueden ofrecer complicaciones simultáneas en varias dimensiones [...]” (García Cela, 2001).

La arbitrariedad del signo es, además, la piedra angular de la lingüística de Saussure y la base de todas sus investigaciones en semiología, en las que se esfuerza por discernir los distintos procesos de comunicación utilizados en la comunicación para convencer e influir. Tomaremos como definición de semiología la propuesta por Martinet: *“En el uso común no lingüístico, los términos signo, rasgo y símbolo no están bien definidos y se utilizan indistintamente para designar fenómenos de órdenes radicalmente diferentes. El denominador común de todos los usos es el hecho de que un elemento A representa a un elemento B o sirve de sustituto de éste; la ciencia que estudia los problemas relativos a la representación se llama semiología o semiótica”* (Martinet, 1975, p. 354).

En la misma línea, Michel Collot, en su obra *La matière émotion*, también intenta describir este vehículo directo e intrínseco que existe entre el hablante y la palabra. Una especie de vínculo tan profundo que resulta estar en consonancia con lo más íntimo de cada persona: *“De cada objeto todos poseemos “una idea profunda” formada por la incesante sedimentación de las “impresiones” que hemos “recibido” de él desde la infancia”* (Collot, 1997, p. 49).

Esta teoría coincide también con la del escritor y filólogo Victor Klemperer, que realizó numerosas investigaciones sobre el lenguaje y más concretamente sobre el del Tercer Reich y denunció, en muchas ocasiones, esta parte del inconsciente ligada al lenguaje: *“Siempre se cita esta frase de Talleyrand, según la cual el lenguaje estaría ahí para ocultar los pensamientos del diplomático. Pero es exactamente lo contrario. Lo que alguien quiere ocultar deliberadamente a los demás o a sí mismo, y también lo que lleva dentro de sí inconscientemente, el lenguaje lo saca a la luz”* (Klemperer, 1996, p. 35).

En definitiva, la palabra saca a la luz una parte del inconsciente del hablante sin que éste sea capaz de darse cuenta. Por eso el lenguaje es tan complejo y extraordinariamente ambiguo. Las palabras están llenas de

sentido y significado, de ahí nuestra perpetua e insaciable búsqueda de la comunicación porque, como señala L'Haire, "[...] *en general, al comunicarse, los seres humanos negocian constantemente el significado, por ejemplo, a través de aclaraciones durante la conversación*" (L'Haire, 2011).

Entonces, ¿cómo abordar esta dificultad, cómo concienciar sobre el uso y las consecuencias de cada palabra en nuestro idioma? Dominar la primera lengua y sus diferentes matices para poder dominar la segunda a su vez, o incluso simultáneamente. Estos son los problemas y retos a los que se enfrentan los filólogos cada día. Por lo tanto, nos pareció importante, en esta tesis, explorar a fondo todos los métodos disponibles para consolidar las enseñanzas de la primera lengua con el fin de transmitir las de manera inteligible y pragmática.

### **1.1.2. Aplicación de una metodología**

En mi contexto profesional, en las distintas clases en las que he podido evolucionar y he visto evolucionar, he tenido la suerte y el placer de constatar que, a lo largo de las últimas décadas<sup>1</sup>, los idiomas y las nuevas tecnologías se han perfeccionado conjuntamente en el ámbito de la enseñanza, lo que ha sido muy favorable tanto para el profesorado como para el alumnado.

Sin embargo, me parece que, aunque estos campos son hoy en día complementarios, no van necesariamente de la mano, al menos no en el día a día de nuestro sector. Por lo tanto, nos pareció interesante y necesario cuestionar la importancia de una posible dualidad entre la lingüística y la informática en el seno de nuestra enseñanza de idiomas.

En esta tesis, examinaremos las humanidades digitales, explotaremos sus mecanismos y automatismos para poder compartirlos e implementarlos en el

---

<sup>1</sup> En teoría, en el contexto francés, las nuevas tecnologías están presentes en la vida de los profesores desde los años 80, gracias sobre todo al plan informático puesto en marcha por el entonces Primer Ministro francés Laurent Fabius. Sin embargo, en la práctica, hace sólo unas décadas que estas herramientas se utilizan realmente en el día a día de las escuelas y universidades.



día a día de profesores y alumnos, porque creemos que percibir sus ventajas en nuestra profesión será sinónimo de avances y mejoras.

Esta tesis se inscribe, por tanto, en esta reflexión y problemática de investigación: ¿puede la ingeniería informática aportar un avance significativo en el ámbito de la enseñanza de idiomas? A lo largo de esta tesis, intentaremos responder a esta cuestión pedagógica y didáctica, explotando estadística y lingüísticamente un corpus importante.

Un corpus importante, sobre todo por su tamaño. Desambiguemos el término "importante" porque aquí debe tomarse en los dos sentidos del término, es decir, tanto en términos de masa voluminosa de información como en términos de contenido memorable. En efecto, la elección de nuestro estudio se basa en los artículos de prensa de los periódicos franceses en el momento de la proclamación de la Gran Guerra. Será un estudio comparativo de cuatro líneas editoriales para contrastar la información y tratar de garantizar la visión de conjunto y la interpretación de los discursos de la época a través de la lectura automática de los mismos.

Por último, precisemos que esta elección del corpus germinó en nuestras mentes cuando notamos, durante las clases y a lo largo del año, y en particular con las generaciones más jóvenes, que nuestros alumnos dudaban u olvidaban las fechas de las dos guerras mundiales o de los principales acontecimientos históricamente cruciales.

Esta observación nos preocupó y por ello pensamos e investigamos en la enseñanza de la lengua sobre un corpus cuyo contenido, tanto lingüístico como histórico, pudiera aportar conocimientos y bases sólidas a los alumnos que comparten un mismo espacio común, que es la Unión Europea. De hecho, se trataba de pensar en un corpus que pudiera garantizar el aprendizaje no sólo a nivel lingüístico, sino también de forma transversal, es decir, a nivel cultural e interpersonal.

Además, numerosos estudios han demostrado que el aprendizaje a partir de un corpus es beneficioso y que, desgraciadamente, todavía se utiliza muy poco en Francia y en todos los países europeos que no son anglosajones.

Como señala Boulton, *"el uso directo de los corpus por parte de los profesores y, sobre todo, de los alumnos no está tan avanzado como sugieren muchos artículos [...] El retraso es aún mayor en Francia, en parte por razones culturales [...]"* (Boulton, 2007). Sin embargo, es cierto que en las últimas décadas se ha abierto el camino hacia el desarrollo de métodos de aprendizaje automático y, por tanto, de forma subyacente, de corpus.

Nuestro proyecto consistirá en implantar más enseñanza basada en corpus auténticos que puedan ser explotados no sólo por el profesor, sino también directamente por los alumnos. Como ya hemos señalado, en el mundo anglosajón estas prácticas son ya habituales y se denominan comúnmente DDL (Data-Driven Learning), que consiste en la explotación directa de los datos de un corpus por parte del alumno.

Afortunadamente, como también señala Molina en su reciente trabajo sobre el Aprendizaje de las Lenguas Asistido por Ordenador (ALAO), *"el trabajo basado en corpus se ha hecho cada vez más popular en el campo de la lingüística y la didáctica de las lenguas en los últimos años"* (Molina, 2017).

Nuestro objetivo será proporcionar un entorno didáctico en el que el tratamiento de corpus pueda ser asistido por máquinas para realizar análisis complejos y precisos de un conjunto de textos de gran tamaño con el fin de extraer los diferentes fenómenos lingüísticos, es decir, todas las locuciones, términos y expresiones fijas de una lengua como la fraseología, o para realizar estudios comparativos. Porque, como explica Chomsky, al poder plantear el lenguaje como un conjunto finito, su procesamiento automatizado es ejecutable: *"Consideramos que un lenguaje es un conjunto (finito o infinito) de oraciones, cada una de ellas de longitud finita y construida por concatenación en un conjunto finito de elementos. Esta definición incluye tanto los lenguajes naturales como los artificiales en la lógica y en las teorías de la programación informática"* (Chomsky, 1971, p. 15).

Nuestra investigación se centrará precisamente en este posible interés que el uso de herramientas de Procesamiento Automático del Lenguaje podría aportar dentro de una clase de FLE para alumnos de nivel intermedio o avanzado (B1, B2, C1, C2). Diferentes análisis automáticos y exhaustivos

sobre los diferentes aspectos morfológicos, sintácticos y semánticos del lenguaje para dar cuenta del "poder" de las palabras y concebir que puedan ser la más noble pero también la más temible de las armas porque la frontera entre la palabra y la ideología puede resultar ser imperceptible.

Definimos el término ideología precisamente en el sentido en que Macherey lo define en su exhaustivo análisis de la ideología y la palabra, es decir, como un término cuya connotación puede ser tanto negativa como positiva:

La ideología es pensamiento en la medida en que se presenta como pensamiento del otro, ya sea este otro el adversario que todo individuo políticamente comprometido combate, ya sea el otro grupo contra el que un grupo determinado define y defiende sus propias opciones, o si es el otro en el sentido más general de la alteridad que la sociedad incluye en su constitución, lo que le prohíbe ser considerada como una totalidad que existe de forma homogénea, y que por lo tanto tiene desde el principio a la manera de un dato primario que no puede ser puesto en cuestión de su propia identidad (Macherey, 2008).

Se trata, pues, de proponer a los alumnos que mejoren sus competencias lingüísticas, pero también sus herramientas lingüísticas, discursivas o persuasivas, para avanzar hacia una reflexión sobre el lenguaje, su precisión, pero también su destreza y su capacidad de seducción y, por tanto, por extensión, de influencia negativa o positiva.

En definitiva, se trata de ofrecer a nuestros alumnos un aprendizaje de la lengua que sea a la vez completo y complejo porque, como ya subrayó Chanier, *"entre los investigadores y los profesionales de la adquisición de la lengua está bien establecido que hablar una lengua con fluidez no puede reducirse a la adquisición de la competencia gramatical. Dado que la lengua y la comunicación son inseparables, se trata de una competencia comunicativa que debe adquirirse"* (Chanier, 1995).

### **1.1.3 La sostenibilidad de las lenguas en la era digital**

En primer lugar, es fundamental recordar que las nuevas tecnologías son actualmente sinónimo de futuro y de progreso, y que se están introduciendo tanto en nuestras universidades como en el conjunto de nuestras carreras, para que sean eficaces y de calidad. Como la sostenibilidad de los idiomas depende ahora de la fusión con la comunicación digital, era urgente adaptar nuestra oferta a esta creciente demanda.

Las humanidades por sí solas ya no pueden pretender tener un futuro seguro; la tecnología de la información es, por tanto, la garantía de una nueva longevidad para ellas. La comunicación digital es ahora el nuevo medio viable y sostenible de nuestra educación humanística porque está en consonancia con los nuevos conceptos de nuestro mundo moderno.

En el contexto de las universidades, el especial interés por las nuevas tecnologías se reflejó también en los medios de comunicación y, especialmente, en la prensa escrita, que insistió en la importancia de promover e incorporar este bien a nuestra vida cotidiana, ya que las ventajas eran numerosas y diversas, al ser el principal vehículo de contacto con el exterior y con los nuevos empleos que implican la informática, Internet y los nuevos conocimientos.

Era también, y sobre todo, el futuro de nuestras universidades al convertirse en nuestro nuevo medio. Las cartas, los idiomas y todos los demás tipos de comunicación han sobrevivido y se han perfeccionado en gran medida gracias al medio digital. Por lo tanto, es esencial dominar las nuevas herramientas de comunicación e intercambio.

A lo largo de las últimas décadas, diversos agentes académicos y económicos han cuestionado la importancia de las nuevas tecnologías. Por ello, es fundamental que se adapten lo antes posible a estos cambios si no quieren perder resultados. Además de los impresionantes avances en las nuevas competencias que se han incorporado rápidamente a nuestro sistema

educativo, también es importante reaccionar lo antes posible y adaptar nuestro sistema educativo a esta nueva globalización.

En la era de Internet, es esencial que nuestra formación humanística acepte e integre plenamente el concepto de informática, ya que es totalmente complementario e indispensable. Cabe destacar que muchas de nuestras universidades han mantenido voluntariamente el perfil humanista que ha caracterizado nuestra tradición y conocimiento a lo largo de los años. De este modo, lo que podría haberse percibido como una contradicción se ha convertido en nuestra baza y nuestro mayor reto.

Recordemos que el dominio del aprendizaje instrumental de la lengua extranjera y sus habilidades didácticas ha mantenido una estrecha relación con una formación primaria en filología hispánica, latín y otras lenguas modernas. Se ha tratado de conservar los estudios de historia cultural, de historia literaria y de historia de la lengua en el núcleo de la formación francesa, por ejemplo, pero también de lingüística sincrónica y diacrónica, junto a ciertas perspectivas comparativas de desarrollos más recientes, como los estudios culturales comparados, la relación entre literatura y cine, la lengua y los medios de comunicación. Por no hablar de la creciente atención al conjunto de la Francofonía.

De hecho, las comisiones que elaboran los planes de estudio han mantenido expresamente este perfil humanista, en el convencimiento de que proporciona la mejor formación lingüística e intelectual para los futuros profesores, traductores, agentes culturales, etc.

Sin embargo, la formación tecnológica es ahora la condición necesaria para que la formación humanística se incorpore y se integre en el mercado laboral actual. Este nuevo activo, además de ser compatible con la formación filológica, es también la condición necesaria para su supervivencia en un mundo en el que el formato y la naturaleza de los medios de conocimiento han evolucionado claramente.

El filólogo, el profesor o el traductor deben, pues, dominar estas nuevas herramientas porque representan una ayuda innegable e indispensable. En

efecto, la programación informática permite realizar muchos análisis muy complejos en unos pocos clics, sin posibilidad de error. Las bases de datos informáticas, por ejemplo, permiten unir varias lenguas para facilitar y perfeccionar las traducciones.

Un buen conocimiento de Internet también permite explotar materiales de enseñanza y aprendizaje más interactivos para alumnos mucho más acostumbrados y receptivos a las nuevas tecnologías, pero también crear vínculos e intercambios internacionales, tanto humanos como comerciales, lo que supone una ventaja considerable para la expansión de las regiones, los territorios y el mundo en general.

Además, puesto que las universidades son cada vez más exigentes, deben estar al tanto del mercado laboral para adaptarse a la oferta de trabajo. En consecuencia, deben aprovechar este auge para formar a sus alumnos en un aspecto profesional y de demanda real en el día a día.

Hoy en día las universidades deben estar en constante sintonía con el mundo exterior, pues de lo contrario se perderán, especialmente en el caso de España, donde la tasa de desempleo está alcanzando niveles alarmantes, mientras que ciertos puestos para nuevos perfiles profesionales (informáticos, expertos en redes sociales, diseñadores web, etc.) no se cubren por falta de personal competente. Estos resultados son extremadamente frustrantes, pero las universidades pueden remediar estas deficiencias. Cabe destacar que las universidades privadas ya han orientado su oferta a cursos con un alto grado de empleabilidad.

En otras palabras, hay que cambiar de actitud en cuanto a la estanqueidad de los planes de estudio, ya que ahora es necesario que la lingüística integre la informática y viceversa. Al igual que la ciencia ha abierto sus puertas a lo abstracto y a lo desconocido para poder cuestionarse y superarse continuamente, ahora se trata de compatibilizar y armonizar los distintos sectores para ser más versátiles y universales. Para ello, y para que la formación filológica sea aplicable y aplicada, se requiere una formación adicional en línea con las herramientas informáticas.

Por ello, los nuevos métodos de enseñanza y análisis de la lengua, cada vez más precisos y eficaces, nos han llevado a profesores e investigadores a las Tecnologías de la Información y la Comunicación (TIC), ya que para todo filólogo representan un punto crucial entre una nueva era tecnológica y la posible perpetuación y explotación de la lengua escrita.

## **1.2. De las TIC al TAL**

### **1.2.1. Tecnología digital y didáctica**

Recordemos que a finales del siglo XX y principios del XXI, lo que permitió que la lengua y la escritura se reinventaran y crecieran aún más fue la fusión entre los ordenadores y los textos. En efecto, son innumerables los textos que se han digitalizado o incluso mecanografiado completa y manualmente, lo que ha permitido, por extensión, archivar y consultar a golpe de ratón toda una serie de documentos y recursos muy valiosos para la investigación y la humanidad en su conjunto.

Este nuevo espacio de aprendizaje, conocido como TIC, que se ha implantado con fuerza en las escuelas y centros educativos desde hace una década, nos lleva ahora de forma natural al TAL (Tratamiento Automático de las Lenguas) y a las múltiples posibilidades que ofrece, ya que es, en cierto modo, la extensión y profundización de estas nuevas tecnologías. Una especie de nueva cohabitación, necesaria e inevitable, entre la lingüística y la informática.

Recordemos la primera definición de TAL dada por Fuchs en 1993:

[...] El procesamiento automático del lenguaje se ocupa de los datos lingüísticos (textos) expresados en un lenguaje (natural), y para poder procesar estos datos de forma automática, es necesario poder explicitar las

reglas del lenguaje, representarlas en formalismos operativos y computables e implementarlas con la ayuda de programas (Fuchs, 1993, p. 13)

Señalemos que, aunque estos dos campos han estado al lado durante mucho tiempo, este reciente entusiasmo por las nuevas tecnologías y, en particular, por la herramienta informática, ha dado lugar a nuevas cuestiones sobre las que varios especialistas se cuestionan, como explica el profesor e investigador Fabrice Issac *"El uso de la informática en un marco lingüístico es muy antiguo [...] El uso masivo de la informática parece crear, de hecho, un cambio en el perfil de ciertos actores de la investigación lingüística: ¿soy informático o lingüista?"* (Issac, 2018: 1). En definitiva, una dualidad de dos entidades contiguas que plantean constantemente nuevos retos y problemas.

Una cuestión tanto más importante cuanto que va a configurar o incluso condicionar la pertinencia de las investigaciones y los resultados para garantizar una interpretación correcta. El lingüista dispone ahora de una herramienta muy poderosa, la de la informática, pero no debe en ningún caso desvirtuar su función principal.

Estas herramientas proporcionan al lingüista numerosas "pistas" de lectura, que luego debe analizar escrupulosamente desde su propio ámbito científico. Como señala J-M Molina, *"hay que admitir que la informática y el TAL pueden ayudar a resolver un buen número de problemas relacionados con la lingüística con porcentajes de éxito bastante significativos, pero siempre será necesario el trabajo del lingüista para verificar la salida de la máquina."* (Molina, 2015, p.54).

A raíz de esto, nos planteamos la siguiente cuestión: ¿Es el TAL una ciencia a caballo entre la informática y la lingüística? Y si es así, ¿dónde debe ubicarse y cómo debe definirse su papel dentro de la enseñanza? El propio profesor Fabrice Issac intenta responder a esta pregunta evocando la función del lingüista computacional: *"El reto es, por tanto, nombrar y describir el modo en que la informática aporta nuevas herramientas y, por tanto, nuevos puntos de vista teóricos sobre la lingüística"* (Issac, 2014: 13).



Es evidente que el TAL es una herramienta extremadamente poderosa para un filólogo apasionado por "diseccionar" un texto para interpretarlo y comprenderlo mejor, ya que el TAL ofrece la posibilidad de realizar análisis complejos y precisos sobre múltiples corpus de gran tamaño.

En un futuro próximo, podría convertirse en una herramienta indispensable para ayudar al filólogo y al profesor en sus tareas cotidianas de análisis del discurso porque los resultados son reveladores, como indica el lingüista Robert Martin, "[...] *el análisis automático obliga a la formalización, es decir, a las formulaciones explícitas y unívocas [...] Sometida a la automatización, la descripción lingüística gana en rigor y a veces en pertinencia*" (Martin, 2001).

El TAL es, por tanto, un poderoso instrumento que se define como "multidisciplinar ya que puede reunir a lingüistas, informáticos, lógicos, psicólogos, documentalistas, lexicólogos y traductores". En resumen, no menos de siete campos están continuamente correlacionados gracias a la Inteligencia Artificial (IA), como señala la Enciclopedia Universalis de Francia.

La Inteligencia Artificial es cada vez más eficaz y ha permitido crear diccionarios electrónicos, establecer traducciones lingüísticas automáticas o semiautomáticas, correctores ortográficos, resúmenes automáticos y hacer posible el acceso inteligente a la información, es decir, al tratamiento automático y rapidísimo de los datos.

En el ámbito del TAL, las primeras investigaciones se centraron en el análisis de las lenguas naturales, en particular la traducción, entre 1949 y 1965. Hay que señalar que este interés se debió principalmente al contexto histórico en el que se produjo, es decir, durante la Guerra Fría en 1950, ya que el principal objetivo de los estadounidenses en aquella época era acceder a todos los contenidos producidos en ruso.

Basándose en las teorías de Shannon, los estadounidenses invirtieron sumas considerables en el desarrollo de la traducción automática porque, como señala Bouillon, era "*una tarea fácil que no debería requerir otras técnicas que las desarrolladas con éxito para descodificar mensajes codificados*" (Bouillon, 1998, p. 9).

Este entusiasmo exponencial por el TAL condujo, a principios de los años 50, al primer congreso internacional sobre traducción automática (TA), organizado por W. Weaver en el centro de IBM en Georgetown en 1952. Esta conferencia fue la primera exposición pública de un sistema automatizado inteligente capaz de traducir del ruso al inglés.

Los años de investigación e inversión en este campo continuaron. Tanto es así que en 1964 el gobierno estadounidense creó un comité especial para evaluar los resultados de la TA. Por desgracia, los resultados fueron decepcionantes y el informe del Comité Asesor de Procesamiento Automático del Lenguaje (Automatic Language Processing Advisory Committee - ALPAC) calificó la traducción automática de utópica. Como dice Kübler, *"este informe concluye que la TA es más lenta, menos eficiente y dos veces más cara que la traducción por humanos"* (Kübler, 2007).

Sin embargo, a finales de los años 50 y 60, las investigaciones del lingüista estadounidense N. Chomsky despertaron nuevas esperanzas en el TAL. Como se recoge en su libro "Language and Thought", publicado en Nueva York en 1968, donde expresa sus impresiones sobre la posición de la lingüística matemática:

La lingüística matemática parece estar hoy en una posición totalmente favorable, entre los enfoques matemáticos de las ciencias sociales y fisiológicas, para desarrollarse no sólo como mera teoría de los hechos, sino también como estudio de los principios y estructuras altamente abstractos que determinan el carácter de los procesos mentales humanos. En este caso, los procesos mentales en cuestión son los que intervienen en la organización de un dominio específico del conocimiento humano, a saber: el conocimiento del lenguaje. (Chomsky, 1968, p.107)

Durante estos años, la experiencia de Chomsky como lingüista marcó un punto de inflexión en el progreso del TAL al proponer, a través de sus libros "Syntactic structures" - 1957 y "Aspects of one Theory of Syntax" - 1965, una teoría del lenguaje basada en dos niveles de representación: la estructura profunda y la estructura superficial.

La estructura profunda según Chomsky se basa, por un lado, en la sintaxis y en la serie de reglas que determinan los sintagmas y otras unidades básicas de la oración, es decir, las reglas de "inserción léxica", que permiten hacer coincidir las palabras del léxico con las categorías gramaticales. La estructura superficial, en cambio, se basa en las reglas de la semántica y las reglas de la fonología. Chomsky define estos dos conceptos de la siguiente manera: "*el componente sintáctico de una gramática debe caracterizar para cada frase una estructura profunda que determina su interpretación semántica y una estructura superficial que determina su interpretación fonética*" (Chomsky, 1971: 31).

Como explica Chomsky con más detalle en su obra fundamental *Lingüística cartesiana*, se trata de discernir la lingüística profunda de la superficial, como si el lenguaje tuviera dos capas: "*Podemos distinguir entre la 'estructura profunda' de una frase y su 'estructura superficial'. La primera es la estructura abstracta básica que determina su interpretación semántica; la segunda, la organización superficial de unidades que determinan la interpretación fonética y que está relacionada con la forma física de la expresión real*" (Chomsky, 1966, p.78).

En su libro *"Syntactic Structures"*, N. Chomsky propone aplicar este proceso de separación dentro de la máquina y disociar estos dos estratos porque es en el nivel de la estructura superficial, es decir, en lo que respecta a la organización superficial de las unidades, donde la máquina es capaz de identificar y reconocer los elementos de un corpus.

Como explica Chomsky, en el caso de que una gramática esté previamente implementada en la máquina, se obtiene un reconocimiento sintáctico de las unidades. Es ahí donde Chomsky desglosa las distintas etapas del reconocimiento automático. En la primera etapa (I), inserta el texto en la máquina (entrada) y luego extrae una gramática del mismo (salida). Esto es lo que él llama el procedimiento de descubrimiento, en el que utiliza consideraciones de probabilidad para distinguir automáticamente las oraciones gramaticales de las no gramaticales. Las llamadas oraciones gramaticales son aquellas que son heterogéneas: "*La diferencia radica en*

*que la noción de 'corrección' abarcaba cosas bastante heterogéneas"* (Ruwet, 1966: 29).

En una segunda etapa, reproduce la misma operación, pero introduce tanto el corpus como la gramática que obtuvo en la primera operación. La máquina reconocerá entonces lo que se le ha enseñado y distinguirá entre las llamadas frases gramaticales con "sí" y las no gramaticales con "no", esto es lo que él llama el procedimiento de decisión.

Por último, introduce en la máquina el corpus y varias gramáticas obtenidas en el procedimiento de decisión a partir de otros corpus para que la máquina elija, entre estas gramáticas, la que sea relativamente mejor, es lo que denomina procedimiento de evaluación.

Citamos aquí el extracto en el que Chomsky define estas diferentes fases:

La figura (36 I), representa una teoría concebida como una máquina con un corpus en la entrada y una gramática en la salida; es por tanto una teoría que proporciona un procedimiento de descubrimiento. (36 II) es una máquina que tiene una gramática y un corpus como entrada y las respuestas "sí" y "no" como salida, indicando si la gramática es la correcta o no; este es el esquema de una teoría que proporciona un procedimiento de decisión. (36 III) representa una teoría con las gramáticas de entrada G1, G2 y el corpus y la gramática de salida G1 o G2 como la mejor: esta teoría proporciona un procedimiento de evaluación. (Chomsky, 1969, p. 57)

Concluimos que Chomsky, al exponer su marco de la teoría generativa, propuso un estudio del proceso lingüístico a través del estudio metódico del sistema sintáctico. Esto ayudó en gran medida al desarrollo del Procesamiento Automático del Lenguaje porque fue la primera vez que se pudo ayudar a la máquina a identificar el significado de las palabras a través del sistema sintáctico. Esto se justifica por el hecho de que Chomsky consideró que era más concebible aislar esta área que otro fenómeno cognitivo:

Es natural [...] suponer que la interpretación semántica de una frase está determinada por el contenido semántico intrínseco de los elementos léxicos, y por la forma en que éstos se relacionan entre sí, en el nivel de la estructura profunda. Asumir esto [...] es decir que son las estructuras profundas las que determinan la interpretación semántica, bajo el efecto de las reglas de interpretación semántica. En esta teoría, pues, se requiere que las estructuras profundas cumplan varias condiciones. En primer lugar, determinan la representación semántica. En segundo lugar, se convierten en estructuras superficiales bien formadas mediante transformaciones gramaticales. Por último, satisfacen el conjunto de condiciones formales definidas por las normas básicas. (Chomsky, 1975, p. 13)

Ese mismo año, 1975, la máquina se convirtió en una herramienta verdaderamente revolucionaria porque la Comunidad Europea, ante una demanda exponencial de traducciones, emprendió la investigación del tratamiento informático y, en particular, de la traducción automática. Por ello, la Comunidad Europea decidió dedicar e invertir una gran cantidad de dinero en este ámbito.

Hoy en día y más que nunca, el tratamiento informático sigue representando, y es importante recordarlo, un mercado en rápida expansión. Digamos que la detección de palabras, gracias a estas diferentes herramientas, tiene un fuerte interés socioeconómico que amasa colosales sumas de dinero puesto que las empresas pueden extraer palabras clave de inmensas bases de datos que luego pueden analizar, controlar, manipular y revender a su antojo.

Estos diferentes elementos nos indican que, en el transcurso del último siglo, la tecnología se desarrolló de forma exponencial, llevando a profesores e investigadores a cuestionar también cada vez más el lugar del TAL en nuestras aulas.

Sin embargo, ¿cómo sacarle el mejor partido a estos recursos? ¿Qué beneficios y a costa de qué inconvenientes? ¿Cuáles son las dificultades y los desafíos que están detrás de estos nuevos métodos de enseñanza?

Todas estas preguntas, y otras más, son a las que hemos pretendido dar respuesta a lo largo de esta tesis. Partiendo de diversos documentos auténticos de la prensa francesa de la Primera Guerra Mundial con tendencias políticas diversas (*L'Humanité, Le Temps, Le Figaro, La Croix*) que tuvimos que limpiar escrupulosamente para nuestra manipulación informática, hemos intentado construir una reflexión en torno al posible aprendizaje de la lengua y de la lingüística desde la perspectiva del TALN.

Con esto en mente, nos hemos planteado estructurar nuestra investigación en diferentes etapas: limpieza de corpus, manejo de la plataforma multidimensional, experimentos lexicométricos y textométricos diversos desde una perspectiva sociocultural y lingüística y, finalmente, una puesta en práctica de nuestra investigación desde un enfoque didáctico con la concepción de ejercicios semi-automatizados.

Así pues, se trata de presentar una propuesta de enseñanza del idioma y de la lingüística bajo el prisma de la inteligencia artificial y del TALN, pero también de la historia y de sus aspectos socioculturales, fuentes vivas y actuales con las que proporcionar una formación general a nuestros alumnos universitarios así como una formación más especializada para otras disciplinas.

Todos ello se puede abordar en diferentes ángulos que han alimentado nuestro constante interés por la investigación. Esperamos, gracias a esta tesis, aportar nuestro granito de arena a la gran labor que es la de la enseñanza de idiomas, buscando siempre una proyección diferente e innovadora.

## **V. CONCLUSIÓN (EVALUACIÓN Y PERSPECTIVAS)**

### **5.1. Resumen del argumento**

Al final de nuestra aproximación, que a veces se asemejaba a una especie de "implacabilidad lexicométrica", pudimos poner en práctica una multitud de funciones de TAL, lo que nos permitió cuestionar y explotar nuestro corpus, que fue para nosotros, a veces, un simple pretexto para la experimentación estadística y las comparaciones técnicas.

Es cierto que este corpus, al que hemos dedicado tanto tiempo, podría haber sido un corpus completamente diferente, pero hemos querido levantar el velo sobre una dificultad que nos parecía ampliamente valiosa y nos permitía, al mismo tiempo, saludar la ardua y minuciosa tarea de ciertos historiadores que consiste en reconstituir los textos, o incluso restaurarlos, como ha sido el caso de nuestro corpus, ya que era imposible lograr acceder a una digitalización adecuada sin recurrir al humano.

Afortunadamente, los informáticos han hecho posible el extraordinario archivo y el intercambio de datos masivos mediante la digitalización, pero esto no siempre es viable para determinados textos.

Por lo tanto, fue una oportunidad para insistir en el hecho de que, si hoy en día tenemos acceso a una gama muy diversa y completa de textos digitalizados: obras literarias, artículos de prensa, etc.; es decir, innumerables soportes que nos ofrecen los TICs y que los profesores de FLE están encantados de poder explotar permanentemente con fines didácticos, es gracias a los avances más o menos recientes de las nuevas tecnologías.

Por ello, nos ha parecido importante aprovechar esta tesis como ejemplo de puesta en práctica de un corpus que no está necesariamente disponible y proponer soluciones a estas importantes dificultades.

Nos pareció una especie de pretexto para rendir homenaje a todo el trabajo de digitalización que se ha realizado en el último siglo y que hace posible y sumamente fácil compartir todos estos extraordinarios contenidos con el único objetivo de hacer accesible el conocimiento.

Para nosotros, era una forma de destacar el arduo trabajo que ha supuesto el ejercicio de digitalización, en algunos casos, incluido el nuestro. Un proceso que, sin embargo, deja un valiosísimo patrimonio y que ha abierto el acceso a un sinfín de textos tan utilizados y tan presentes en nuestra realidad cotidiana y, por supuesto, más que nunca en nuestra enseñanza actual.

Es importante recordar que consideramos que estos materiales son la garantía de un proceso de aprendizaje transversal, rico, multidisciplinar y, sobre todo, accesible.

Este trabajo nos ha obligado a veces a tomar caminos muy sinuosos en los que hemos tenido que redoblar nuestros esfuerzos para poder avanzar. Incluso planteó cuestiones de legitimidad: como profesor de francés como lengua extranjera, ¿tengo legitimidad para abordar campos tan remotos como la informática, la historia o la sociolingüística? Todas estas cuestiones nos obligaron a empujar la puerta de otras disciplinas y nos animaron a crear fuertes relaciones entre estas diferentes ciencias, inscribiendo así nuestra tesis dentro de las humanidades digitales.

Relaciones que, por supuesto, son muy recomendables ya que estamos evolucionando en un siglo donde la tecnología nos ha permitido perfeccionar nuestra enseñanza haciéndola más accesible y más parecida a los medios a los que están acostumbrados nuestros alumnos.

## **5.2. Resolución de la cuestión**

En esta tesis, nos hemos visto abocados a reflexionar sobre dos cuestiones, una llevada a la otra. En efecto, nuestra principal



problemática se refería al interés de la aplicación de las herramientas de TAL en el marco de nuestra enseñanza y la siguiente, una problemática directamente vinculada a nuestro corpus para aprovecharlo al máximo.

Por ello, formulamos las siguientes preguntas: por un lado, ¿es la ingeniería informática capaz de aportar un avance significativo en el ámbito de la enseñanza de idiomas? Y, en segundo lugar: ¿sería posible, a través del léxico utilizado en cada uno de los artículos, demostrar la adhesión o el rechazo de las diferentes líneas editoriales en relación con el destino de Europa y del mundo en la entrada de la Primera Guerra Mundial?

Una vez concluida esta tesis, estamos en condiciones de dar respuestas precisas a las distintas preguntas que surgieron al principio de nuestra reflexión.

Para ello, es conveniente hacer un repaso de las múltiples operaciones computacionales que hemos realizado, ya que nos permiten afirmar en esta conclusión que las construcciones de partes, subpartes y particiones, entre otras, han sido elementos clave para la realización de análisis contrastivos avanzados.

Los abundantes cálculos sobre las propiedades de palabras como sustantivos, adjetivos, verbos u otras categorías gramaticales han puesto de manifiesto, por su parte, elementos notables en el corpus, como sus concordancias, co-ocurrencias o relaciones fraseológicas.

Estas correspondencias nos llevaron a reflexionar sobre las nociones de palabras con mayor frecuencia, predicado, fijación, semi-fijación, secuencia libre, analogía, sinonimia, antonimia, polisemia y relaciones concesivas.

Por último, y para ir más allá, las distintas funciones, como los análisis de correspondencia factorial o los gráficos de densidad, nos ofrecieron la posibilidad de una visión totalmente original en cuanto al enfoque de determinadas palabras.

Tras las diversas aplicaciones de esta visión general de las herramientas de TAL, podemos concluir que la ingeniería informática es plenamente capaz

de aportar avances significativos en el ámbito de la enseñanza de las lenguas, ya que nuestros alumnos tendrán la posibilidad de adentrarse en la complejidad de la lengua, es decir, en la gramática, la sintaxis o el léxico, a través de numerosas funciones, cálculos y aplicaciones que les permitirán formular hipótesis, verificarlas, pero también y sobre todo comprenderlas.

Un conjunto de elementos que les permitirá desarrollar mecanismos de interpretación para ensayos largos y desarrollar, de este modo, su pensamiento crítico. Podrán incluso cuestionarse sobre la evolución de la lengua y su sintaxis gracias a los análisis diacrónicos y prever la realización de investigaciones en este sentido.

Además, el TAL permite, como vimos en el cuarto capítulo, el diseño de ejercicios semiautomatizados. Esto supone una ventaja considerable para nosotros como profesores, pero también para nuestros alumnos, ya que nos permitirá crear ejercicios personalizados y explotar al máximo nuestro corpus en muy poco tiempo.

Además, las dificultades relacionadas con los corpus también nos han permitido adoptar técnicas robustas para poder trabajar a diario con corpus muy grandes e incluso raros, a los que podríamos haber renunciado si no hubiéramos adquirido mecanismos y conocimientos en este campo.

A raíz de esta observación, podemos concluir, sin equivocarnos, que el TAL contribuye a un progreso significativo en el ámbito de la enseñanza.

En cuanto a la otra cuestión, es decir, si el léxico utilizado en cada uno de los artículos era indicativo de la adhesión o el rechazo de diferentes líneas editoriales en la entrada del conflicto de la PGM, puede decirse que esta tesis ha identificado las estadísticas de las palabras así como sus contextos fraseológicos y que ello ha orientado nuestras intuiciones hacia determinados caminos. Sin embargo, los números no pueden tomarse como verdades absolutas y hemos tenido que volver al texto constantemente para llegar a interpretaciones plausibles.

Los experimentos lexicométricos y textométricos realizados y los diversos resultados obtenidos demostraron que el léxico era decisivo y contribuía en

gran medida a la retórica y la elocuencia del discurso periodístico de cada uno de los periódicos.

Un ejemplo es la palabra "guerra", que ha sido nuestra piedra angular y en torno a la cual hemos realizado una serie de cálculos exhaustivos y reveladores que nos han llevado a diferentes conclusiones, unas veces claras y otras nebulosas, pero que han revelado todas las dimensiones posibles de la palabra.

Las estadísticas revelan ciertas tendencias y pueden desvelar grandes verdades, pero está claro que el discurso puede ocultar otras. Tomemos un ejemplo muy revelador. Muchos historiadores han examinado con detalle y minuciosidad las causas de la PGM, se han movilizado muchas obras y recursos en esta dirección y, por supuesto, muchos investigadores han planteado múltiples hipótesis, pero sólo una ha creado unanimidad. Todos coinciden en que el "miedo" fue EL factor determinante. Una palabra tan pequeña para una catástrofe humana tan grande. Y, sin embargo, es una palabra con una frecuencia insignificante, porque es una palabra que no "habla" estadísticamente.

Entonces, cómo entender que, al declararse la PGM, el discurso periodístico que proclama esta guerra impulsada por el miedo apenas utilice esta palabra. La experiencia nos dice que es porque el miedo no se dice, se insinúa, y en eso consiste el trabajo del filólogo: en leer entre líneas.

Y por eso nuestra relación con el TAL debe basarse en métodos híbridos que combinen continuamente la tecnología con el ser humano. Sin embargo, el potencial del TAL no es despreciable, ya que es extremadamente robusta a la hora de analizar grandes corpus, a diferencia de los humanos, y nos permite, además de un enfoque muy preciso, un enfoque más objetivo. Se trata de un análisis léxico que, en principio, pretende estar "desvinculado" de su contexto. Un contexto que puede, eventualmente, "ahogar" la redundancia de ciertas palabras que luego se instalan discretamente en nuestro inconsciente.

Así, confrontamos diferentes enfoques multidimensionales que nos llevaron a que los tres factores siguientes: la línea editorial, la elección del léxico y la frecuencia, eran susceptibles de poner en marcha una especie de retención ideológica sobre el lector de manera inconsciente. Esto nos permitió concluir que la inevitable subjetividad que existe en el periodismo esclaviza el lenguaje y, por extensión, al lector, ya que el léxico utilizado, además de ser repetitivo, conlleva un fuerte y ambiguo "contenido emocional".

Además, cabe destacar que las herramientas de TAL son capaces de identificar, con un nivel de relevancia bastante alto, numerosos datos que nos permitirán interpretar la obra en su conjunto, ya que existen muchos puntos de convergencia entre la lectura automática y la lectura manual.

Por supuesto, también hay puntos de divergencia, sobre todo en cuanto a los detalles, porque la máquina extraerá principalmente frecuencias y nosotros siempre tendremos que recurrir al texto en su conjunto para detectar ciertas particularidades o sutilezas.

Sin embargo, en un contexto de enseñanza de la lengua o la literatura, este primer enfoque totalmente diferente puede ser muy estimulante para los alumnos porque tendrán que cuestionar los numerosos elementos obtenidos por la máquina y, por supuesto, al estar en la era digital, cuestionar la información es una de las principales perspectivas de nuestra enseñanza.

En resumen, y para intentar reunir las principales conclusiones, ahora estamos en condiciones de decir que el TAL es una ciencia extremadamente densa que ofrece todo un abanico de posibilidades que serán de gran utilidad en la enseñanza de idiomas.

### **5.3. Posición y contribución**

Siguiendo estas diferentes conclusiones, creemos que el TAL tiene su lugar en la enseñanza de idiomas porque contribuye, en gran medida, al

dominio y la comprensión del léxico y la morfosintaxis, tanto en el contenido como en la forma.

Esta tesis nos ha permitido poner en práctica toda una serie de investigaciones y experimentos lexicométricos y textométricos tanto en lo que respecta a la enseñanza de la lengua como a la ideología periodística. Podemos concluir que se trata de una contribución original en el sentido de que esta tesis ha reunido diferentes disciplinas como el lenguaje, la historia, la informática y la sociopolítica.

#### **5.4. Limitaciones**

Sin embargo, esta tesis tiene algunas limitaciones. En efecto, aunque nuestra ambición sea implantar el TAL en el día a día de los profesores, estos métodos no se impondrán de la noche a la mañana, ya que estos programas de Procesamiento Automático del Lenguaje siguen requiriendo un verdadero esfuerzo e interés por parte del profesorado, ya que las diferentes consultas o funciones exigen una cierta habilidad informática y un entusiasmo palpable por este tipo de programas y razonamientos.

Otra limitación es que no hemos podido profundizar en algunas de las investigaciones, especialmente en el ámbito de los AFC o el estudio sincrónico, ya que esto sería objeto de otra tesis.

Además, para futuras tesis, podría ser muy interesante ampliar el campo cronológico de los periódicos y seguir haciendo análisis sincrónicos o diacrónicos sobre todas las fechas que marcan el periodo de conflicto de la PGM o posiblemente relacionadas con los otros grandes acontecimientos del siglo XX mencionados al principio de esta tesis.

La idea es siempre reutilizar la metodología que hemos desarrollado para recoger y procesar la información, y luego sacarle el máximo partido utilizando herramientas de TAL.

## **5.5. Perspectivas y profundización**

Ni que decir tiene que esta tesis nos ha dado la oportunidad de explotar y profundizar en los múltiples recursos y funciones del TAL. Nuestra ambición futura sería poder crear, dentro de nuestro Departamento de lengua francesa, un laboratorio en el que se pudiera acceder a numerosos programas y software de TAL y de reconocimiento óptico de caracteres (OCR) para poder importar numerosos corpus de todo tipo con el fin de realizar diversos cálculos en el marco de la enseñanza, pero también de la investigación, ya que estos diversos métodos de exploración serían una forma constructiva y enriquecedora para que nuestros alumnos conciban el aprendizaje de la lengua y puedan detectar las tendencias de uso del francés.

Además, las comisiones que elaboran los planes de estudio han mantenido expresamente este perfil humanista, en el convencimiento de que proporciona la mejor formación lingüística e intelectual a los futuros profesores, traductores, agentes culturales, etc.

Sin embargo, la formación tecnológica es ahora la condición necesaria para que la formación humanística se incorpore y se integre en el mercado laboral actual. Este nuevo activo, además de ser compatible con la formación filológica, es también la condición necesaria para su supervivencia en un mundo en el que el formato y la naturaleza de los medios de conocimiento han evolucionado claramente.