



MÁSTER EN LETRAS DIGITALES

TRABAJO DE FIN DE MÁSTER

CURSO 2017-2018

**PROYECTO PARA UN INDEXADOR DE CORPUS DE
APRENDICES DE ESPAÑOL**

ESPECIALIDAD: COLECCIONES DIGITALES

APELLIDOS Y NOMBRE: DÍAZ SÁNCHEZ, ALICIA

DNI: 47515707-V

CONVOCATORIA: NOVIEMBRE

TUTORA: ANA FERNÁNDEZ-PAMPILLÓN CESTEROS

Departamento de Filología Románica, Filología Eslava y Lingüística General

MÁSTER EN: Letras Digitales: Estudios Avanzados en Textualidades Electrónicas

COMPROMISO DEONTOLÓGICO PARA LA ELABORACIÓN, REDACCIÓN Y POSIBLE PUBLICACIÓN DEL TRABAJO DE FIN DE MÁSTER (TFM)

CENTRO: Facultad de Filología y Facultad de Informática

ESTUDIANTE DE MÁSTER: Alicia Díaz Sánchez

TUTOR/ES DEL TFM: Ana Fernández-Pampillón Cesteros

TÍTULO DEL TFM: *Proyecto para un indexador de corpus de aprendices de español*

FECHA DE PRIMERA MATRÍCULA: Septiembre de 2017

FECHA DE SEGUNDA MATRÍCULA (en caso de producirse):

1. Objeto

El presente documento constituye un compromiso entre el estudiante matriculado en el Máster en Letras Digitales y su Tutor/es y en el que se fijan las funciones de supervisión del citado trabajo de fin de máster (TFM), los derechos y obligaciones del estudiante y de su/s profesor/es tutor/es del TFM y en donde se especifican el procedimiento de resolución de potenciales conflictos, así como los aspectos relativos a los derechos de propiedad intelectual o industrial que se puedan generar durante el desarrollo de su TFM.

2. Colaboración mutua

El/los tutor/es del TFM y el autor del mismo, en el ámbito de las funciones que a cada uno corresponden, se comprometen a establecer unas condiciones de colaboración que permitan la realización de este trabajo y, finalmente, su defensa de acuerdo con los procedimientos y los plazos que estén establecidos al respecto en la normativa vigente.

3. Normativa

Los firmantes del presente compromiso declaran conocer la normativa vigente reguladora para la realización y defensa de los TFM y aceptan las disposiciones contenidas en la misma.

4. Obligaciones del estudiante de Máster

- Elaborar, consensado con el/los Tutor/es del TFM un cronograma detallado de trabajo que abarque el tiempo total de realización del mismo hasta su lectura.
- Informar regularmente al Tutor/es del TFM de la evolución de su trabajo, los problemas que se le planteen durante su desarrollo y los resultados obtenidos.
- Seguir las indicaciones que, sobre la realización y seguimiento de las actividades formativas y la labor de investigación, le hagan su tutor/es del TFM.
- Velar por el correcto uso de las instalaciones y materiales que se le faciliten por parte de la Universidad Complutense con el objeto de llevar a cabo su actividad de trabajo, estudio e investigación.

5. Obligaciones del tutor/es del TFM

- Supervisar las actividades formativas que desarrolle el estudiante; así como desempeñar todas las funciones que le sean propias, desde el momento de la aceptación de la tutorización hasta su defensa pública.
- Facilitar al estudiante la orientación y el asesoramiento que necesite.

6. Buenas prácticas

El estudiante y el tutor/es del TFM se comprometen a seguir, en todo momento, prácticas de trabajo seguras, conforme a la legislación actual, incluida la adopción de medidas necesarias en materia de salud, seguridad y prevención de riesgos laborales.

También se comprometen a evitar la copia total o parcial no autorizada de una obra ajena presentándola como propia tanto en el TFM como en las obras o los documentos literarios, científicos o artísticos que se generen como resultado del mismo. Para tal, el estudiante firmará la Declaración de No Plagio del ANEXO I, que será incluido como primera página de su TFM.

7. Procedimiento de resolución de conflictos académicos

En el caso de producirse algún conflicto derivado del incumplimiento de alguno de los extremos a los que se extiende el presente compromiso a lo largo del desarrollo de su TFM, incluyéndose la posibilidad de modificación del nombramiento del tutor/es, la coordinación del máster buscará una solución consensuada que pueda ser aceptada por las partes en conflicto. En ningún caso el estudiante podrá cambiar de Tutor directamente sin informar a su antiguo Tutor y sin solicitarlo oficialmente a la Coordinación del Máster.

En el caso de que el conflicto persista se gestionará según lo previsto en el SGIC de la memoria verificada.

8. Confidencialidad

El estudiante que desarrolla un TFM dentro de un Grupo de Investigación de la Universidad Complutense, o en una investigación propia del Tutor, que tenga ya una trayectoria demostrada, o utilizando datos de una empresa/organismo o entidad ajenos a la Universidad Complutense de Madrid, se compromete a mantener en secreto todos los datos e informaciones de carácter confidencial que el Tutor/es del TFM o de cualquier otro miembro del equipo investigador en que esté integrado le proporcionen así como a emplear la información obtenida, exclusivamente, en la realización de su TFM.

Asimismo, el estudiante no revelará ni transferirá a terceros, ni siquiera en los casos de cambio en la tutela del TFM, información del trabajo, ni materiales producto de la investigación, propia o del grupo, en que haya participado sin haber obtenido, de forma expresa y por escrito, la autorización correspondiente del anterior Tutor del TFM.

9. Propiedad intelectual e industrial

Cuando la aportación pueda ser considerada original o sustancial el estudiante que ha elaborado el TFM será reconocido como cotitular de los derechos de propiedad intelectual o industrial que le pudieran corresponder de acuerdo con la legislación vigente.

10. Periodo de Vigencia

Este compromiso entrará en vigor en el momento de su firma y finalizará por alguno de los siguientes supuestos:

- Cuando el estudiante haya defendido su TFM.
- Cuando el estudiante sea dado de baja en el Máster en el que fue admitido.
- Cuando el estudiante haya presentado renuncia escrita a continuar su TFM.
- En caso de incumplimiento de alguna de las cláusulas previstas en el presente documento o en la normativa reguladora de los Estudios de Posgrado de la Universidad Complutense.

ANEXO I: DECLARACIÓN DE NO PLAGIO

D./Dña. Alicia Díaz Sánchez con NIF 47518707V, estudiante de Máster en la Facultad de Filología y en la Facultad de Informática de la Universidad Complutense de Madrid en el curso 2017-2018, como autor/a del trabajo de fin de máster titulado *Proyecto para un indexador de corpus de aprendices de español*

y presentado para la obtención del título correspondiente, cuya tutora es:

Ana Fernández-Pampillón Cesteros

DECLARO QUE:

El trabajo de fin de máster que presento está elaborado por mí y es original. No copio, ni utilizo ideas, formulaciones, citas integrales e ilustraciones de cualquier obra, artículo, memoria, o documento (en versión impresa o electrónica), sin mencionar de forma clara y estricta su origen, tanto en el cuerpo del texto como en la bibliografía. Así mismo declaro que los datos son veraces y que no he hecho uso de información no autorizada de cualquier fuente escrita de otra persona o de cualquier otra fuente.

De igual manera, soy plenamente consciente de que el hecho de no respetar estos extremos es objeto de sanciones universitarias y/o de otro orden.

En Madrid, a 05 de noviembre de 2018

Fdo.: Alicia Díaz Sánchez

Esta DECLARACIÓN debe ser insertada en primera página de todos los trabajos fin de máster conducentes a la obtención del Título.

El patrimonio digital consiste en recursos únicos que son fruto del saber o la expresión de los seres humanos. [...] Los objetos digitales [...] a menudo son efímeros, y su conservación requiere un trabajo específico en este sentido en los procesos de producción, mantenimiento y gestión.

Muchos de esos recursos revisten valor e importancia duraderos, y constituyen por ello un patrimonio digno de protección y conservación en beneficio de las generaciones actuales y futuras. Este legado en constante aumento puede existir en cualquier lengua, cualquier lugar del mundo y cualquier campo de la expresión o el saber humanos.

Artículo 1 de la *Carta sobre preservación del patrimonio digital* de la UNESCO

Resumen

En las últimas décadas se han generado en el ámbito académico multitud de corpus de aprendices de lenguas meta y, por lo tanto, de aprendices de español. Sin embargo, el acceso a estos recursos no está organizado de tal manera que se garantice su localización y recuperación. La propuesta de este proyecto es precisamente la creación de un sistema de información que asegure a investigadores, profesores y estudiantes relacionados con ELE el acceso a los corpus de aprendientes de español. Se defiende como solución más adecuada para centralizar la información dispersa el desarrollo de un indexador (repositorio) con la aplicación OdA del grupo de investigación UCM-ILSA. Esta hipótesis ha sido validada empíricamente a través del diseño, implementación y evaluación de un prototipo del indexador y de la discusión de los resultados. Se puede acceder al sistema mediante el enlace http://repositorios.fdi.ucm.es/corpus_aprendices_español.

Palabras clave

Corpus de aprendices, ELE, repositorios digitales, objetos digitales, colecciones de objetos digitales

Abstract

In the last decades a multitude of learner corpora have been generated in the academic field. However, the access to these resources is not organized in order to guarantee its retrieval. The proposal of this project is the creation of a system which assures the corpora access to the academic community. Concretely, we consider an indexer (repository) the best solution. This hypothesis was empirically verified through design, implementation and evaluation of an indexer's prototype. The system is accessible by the link http://repositorios.fdi.ucm.es/corpus_aprendices_español.

Keywords

Learner corpora, ELE, digital repository, digital object, digital collection

Résumé

Ces dernières décennies, une multitude de corpus d'apprenants de langues cible (et donc d'espagnol) ont été créés dans le domaine académique. Cependant l'accès aux ressources n'est pas organisé de manière que leur localisation et récupération soient garanties. La proposition de ce projet est précisément la création d'un système qui assure l'accès aux corpus d'apprenants d'espagnol à la communauté académique. On défend le développement d'un indexeur comme la solution la plus appropriée. Cette hypothèse a été empiriquement vérifiée grâce au design, à la mise en œuvre et à l'évaluation d'un prototype de l'indexeur. On peut accéder au système à travers le lien http://repositorios.fdi.ucm.es/corpus_aprendices_español.

Mots clés

Corpus d'apprenants, ELE, dépôt numérique, objets numérique, collections d'objets numériques

Índice

Resumen.....	9
Palabras clave	9
Abstract	9
Keywords	9
Résumé.....	9
Mots clés	9
Agradecimientos.....	15
1 Introducción	17
2 Hipótesis de partida y objetivos.....	18
3 Metodología	19
3.1 Fase de análisis.....	19
3.2 Fase de diseño.....	19
3.3 Fase de implementación	19
3.4 Fase de pruebas	19
3.5 Fase de mejoras	19
3.6 Justificación del proyecto.....	20
3.7 Medios materiales.....	20
4 Revisión de antecedentes	20
4.1 Modelos de documentación de corpus y publicaciones.....	23
4.2 Indexador	24
4.3 Oda.....	25
4.3.1 Colecciones de objetos digitales	26
4.3.2 Facetas jerarquizadas.....	26
5 Resultados del proyecto y discusión	27
5.1 Resultados del análisis	27
5.1.1 Especificación de requisitos y revisión de antecedentes	27
5.1.2 Discusión	28
5.2 Resultados del diseño	28
5.2.1 Modelo de metadatos.....	28
5.2.2 Modelo de catalogación.....	33
5.2.3 Discusión	34
5.3 Resultado de la implementación.....	37
5.3.1 El Indexador ICAE	37
5.3.2 Discusión	42
5.4 Resultados de la evaluación	44

5.4.1	Evaluación deductiva de calidad	44
5.4.2	Evaluación empírica de la usabilidad	45
5.4.3	Discusión	51
5.5	Resultados de los cambios y mejoras.....	53
5.5.1	Modificaciones en los títulos de la barra de navegación	53
5.5.2	Modificaciones en las recomendaciones de uso.....	53
5.5.3	Discusión	54
6	Conclusiones y trabajo futuro	55
7	Bibliografía	57
	Anexo I: Diagrama de Gantt	61
	Anexo II: Especificación de requisitos	62
	Anexo III: Modelo entidad-relación	70
	Anexo IV: diagrama UML.....	71
	Anexo V: Cuestionario de evaluación empírica.....	72
	Anexo VI: Resultados de la evaluación empírica.....	76

Índice de ilustraciones

<i>ILUSTRACIÓN 1: TABLA DE DATOS LEARNER CORPORA AROUND THE WORLD</i>	<i>21</i>
<i>ILUSTRACIÓN 2: PRESENTACIÓN DE ICAE.....</i>	<i>37</i>
<i>ILUSTRACIÓN 3: NAVEGACIÓN CLASIFICADA DE LA COLECCIÓN DE CORPUS.....</i>	<i>38</i>
<i>ILUSTRACIÓN 4: NAVEGACIÓN CLASIFICADA DE LA COLECCIÓN DE PUBLICACIONES.....</i>	<i>38</i>
<i>ILUSTRACIÓN 5: APARTADO COLECCIONES</i>	<i>39</i>
<i>ILUSTRACIÓN 6: MUESTRA PARCIAL DEL APARTADO BÚSQUEDA.....</i>	<i>39</i>
<i>ILUSTRACIÓN 7: PESTAÑA DE DATOS DE UN OBJETO DIGITAL DE ICAE</i>	<i>40</i>
<i>ILUSTRACIÓN 8: PESTAÑA DE RECURSOS DE UN OBJETO DIGITAL DE ICAE</i>	<i>41</i>
<i>ILUSTRACIÓN 9: PESTAÑA DE METADATOS DE UN OBJETO DIGITAL DE ICAE</i>	<i>41</i>
<i>ILUSTRACIÓN 10: PÁGINA DE CONTACTO DE ICAE</i>	<i>42</i>
<i>ILUSTRACIÓN 11: PÁGINA DE RECOMENDACIONES DE USO DE ICAE</i>	<i>42</i>
<i>ILUSTRACIÓN 12: NUEVA BARRA DE NAVEGACIÓN</i>	<i>53</i>
<i>ILUSTRACIÓN 13: NUEVA SECCIÓN DE RECOMENDACIONES DE USO</i>	<i>53</i>
<i>ILUSTRACIÓN 14: NUEVO OBJETO DIGITAL DE AYUDA</i>	<i>54</i>
<i>ILUSTRACIÓN 15: NUEVO OBJETO DIGITAL DE AYUDA</i>	<i>54</i>

Índice de tablas

<i>TABLA 1: METADATOS COMUNES A TODAS LAS COLECCIONES.....</i>	<i>28</i>
<i>TABLA 2: METADATOS DE CORPUS.....</i>	<i>30</i>
<i>TABLA 3: METADATOS DE INFORMANTES DEL CORPUS.....</i>	<i>31</i>
<i>TABLA 4: METADATOS DE LENGUAS DE LOS INFORMANTES</i>	<i>31</i>
<i>TABLA 5: METADATOS DE CREADORES.....</i>	<i>31</i>
<i>TABLA 6: METADATOS DE PUBLICACIONES.....</i>	<i>32</i>
<i>TABLA 7: MODELO DE CATALOGACIÓN</i>	<i>33</i>
<i>TABLA 8: MODELO DE CATALOGACIÓN</i>	<i>34</i>

TABLA 9: RESULTADOS DE LA EVALUACIÓN DE CALIDAD.....	45
TABLA 10: RESULTADOS DEL CUESTIONARIO DE USABILIDAD SUS DEL GRUPO 1.....	76
TABLA 11: RESULTADOS DEL CUESTIONARIO DE USABILIDAD SUS DEL GRUPO 2.....	77
TABLA 12: RESULTADOS A LAS SITUACIONES DEL CUESTIONARIO (GRUPO 1).....	79
TABLA 13: RESULTADOS A LAS SITUACIONES DEL CUESTIONARIO (GRUPO 2).....	81
TABLA 14: RESULTADOS DE LAS AFIRMACIONES SOBRE LA UTILIDAD DE LA TERMINOLOGÍA Y LOS APARTADOS (GRUPO 1)	83
TABLA 15: RESULTADOS DE LAS AFIRMACIONES SOBRE LA UTILIDAD DE LA TERMINOLOGÍA Y LOS APARTADOS (GRUPO 2)	84
TABLA 16: RESULTADOS DE ANÁLISIS CUALITATIVO (GRUPO 1).....	92
TABLA 17: RESULTADOS DEL ANÁLISIS CUALITATIVO (GRUPO 2).....	105

Índice de gráficos

GRÁFICO 1: RESULTADOS DE LA ENCUESTA SOBRE LA FACILIDAD DE LAS BÚSQUEDAS PROPUESTAS (GRUPO 1)	47
GRÁFICO 2: RESULTADOS DE LA ENCUESTA SOBRE LA FACILIDAD DE LAS BÚSQUEDAS PROPUESTAS (GRUPO 2)	47
GRÁFICO 3: ENCUESTA SOBRE LA PERTINENCIA DE LA TERMINOLOGÍA DE ICAE (GRUPO 1).....	48
GRÁFICO 4: ENCUESTA SOBRE LA PERTINENCIA DE LA TERMINOLOGÍA DE ICAE (GRUPO 2).....	48
GRÁFICO 5: ENCUESTA SOBRE LA UTILIDAD DE LOS APARTADOS DE ICAE (GRUPO 1)	49
GRÁFICO 6: ENCUESTA SOBRE LA UTILIDAD DE LOS APARTADOS DE ICAE (GRUPO 2)	49
GRÁFICO 7: RESULTADOS DEL ANÁLISIS CUALITATIVO (GRUPO 1)	50
GRÁFICO 8: RESULTADOS DEL ANÁLISIS CUALITATIVO (GRUPO 2)	51

Agradecimientos

A Ana Fernández-Pampillón, por la tutorización del trabajo y la disposición permanente a ayudar.

A José Amenós Pons, por la tutorización de la beca de colaboración y por las ideas.

A Luis García Fernández, por la ayuda y la amabilidad.

Al grupo UCM-ILSA, por el servidor web de aplicaciones.

A los alumnos de las asignaturas Lingüística aplicada a la enseñanza de ELE y Contraste gramatical y cultural en ELE del profesor José Amenós, por haber sido los primeros usuarios de ICAE y haber dado cuenta de ello.

A mis compañeras y a mi madre, por hacer la comida, pasear con Bambú y fregar los suelos.

1 Introducción

En un contexto en el que la información que alberga internet se multiplica a una velocidad incontrolable y una de las grandes preocupaciones es cómo aprehender y mantener todos los datos, información y conocimiento que existen en el mundo para poder utilizarlos, una buena organización de la información que permita recuperarla se vuelve esencial. El problema que se plantea resolver este proyecto es precisamente el de la dispersión y poca visibilidad de un tipo de recursos de gran utilidad para el estudio lingüístico y la enseñanza de lenguas: los corpus de aprendices de lenguas. En concreto, los corpus de aprendices del español como lengua meta.

Los corpus de aprendientes de español, generados por diferentes estudiantes no nativos de español a lo largo y ancho del planeta y recogidos por investigadores y profesores de español como lengua extranjera (en adelante ELE), son recursos útiles tanto para el estudio de la interlengua y la investigación en adquisición de segundas lenguas como para la docencia de estas. Sin embargo, se encuentran tan dispersos por internet que su localización y acceso es complicado. Son necesarias costosas búsquedas en varios idiomas para recuperarlos. Se trata, por tanto, de un inconveniente para los creadores de los corpus, que no sacan todo el partido posible a su trabajo, y para los usuarios, que están pasando por alto materiales que les podrían ser de utilidad.

La dificultad de recuperación, además de provocar que estos materiales permanezcan invisibles, implica el peligro de generar recursos similares por desconocimiento de lo que ya existe. Abordar este problema visibilizando los corpus y poniendo a disposición de la comunidad académica información organizada supone, por tanto, minimizar esos riesgos.

El motor principal de este trabajo es el de generar un sistema doblemente útil para investigadores, estudiantes y profesionales que se ha denominado *Indexador de Corpus de Aprendices de Español*. El Indexador, por un lado, servirá de ayuda a los creadores para dar visibilidad a sus recursos; por otro, ayudará a los potenciales usuarios a recuperarlos con facilidad. El proyecto surge en el marco de una beca de colaboración en el Departamento de Lengua Española y Teoría de la Literatura y Literatura Comparada concedida por la Universidad Complutense de Madrid.

Esta memoria de Trabajo de Fin de Máster presenta y describe el proyecto de construcción del Indexador de Corpus de Aprendices de Español (en adelante ICAE). Se ha organizado en ocho secciones. En la sección 1, se ha hecho una introducción explicando el problema que se pretende abordar, la motivación que ha llevado a querer resolverlo y la estructura que se va a seguir para justificar y documentar el proyecto.

En la sección 2, se explica la hipótesis que se va a defender y los objetivos generales que persigue el proyecto, así como los objetivos específicos que se plantean para lograr el propósito general. Cada objetivo específico se corresponde con una fase del método, que se explica en la sección 3, «Metodología». En este apartado también se señalan los medios materiales con los que ha contado el proyecto.

En la sección 4, se presentan los antecedentes, una revisión de los sistemas que abordan solucionar problemas similares al de este proyecto, de sus aportaciones y de las aportaciones que se pretenden con ICAE así como una breve explicación de las bases teóricas del recurso.

En la sección 5, se detallan y discuten los resultados obtenidos en cada fase del proyecto.

En la sección 6, se explican las conclusiones a las que se ha llegado y se apuntan líneas de trabajo futuras entorno al Indexador de Corpus de Aprendices de Español.

Por último, en la sección 7, se listan las referencias bibliográficas que se han consultado a lo largo del trabajo.

La octava sección la forman los anexos. En ellos se recoge la documentación del proyecto y los resultados de la evaluación: diagrama de organización del trabajo (anexo I), especificación de requisitos del sistema (anexo II), diagrama entidad-relación (anexo III), diagrama UML (anexo IV), cuestionario de la evaluación empírica (anexo V) y resultados de la evaluación empírica (anexo VI).

2 Hipótesis de partida y objetivos

El objetivo principal de este trabajo es resolver el problema generado por la dispersión de materiales e información de los corpus mencionados.

Para lograr este objetivo, la hipótesis que se defiende es que, de todas las formas en que se podría recoger y ordenar la información relativa a corpus de aprendientes de español, una de las más adecuadas es un repositorio o indexador¹. La suposición se fundamenta en las características de este tipo de sistema, que facilitan la creación y mantenimiento de los recursos, así como en la información y ventajas que aportan para la organización de la información y la representación del conocimiento, esenciales para asegurar la localización y selección de los recursos buscados. Estos aspectos se desarrollarán en los antecedentes.

Este trabajo está encaminado, pues, a comprobar la certeza de la hipótesis planteada. Para ello se proponen como demostración empírica de su validez el diseño e implementación de un indexador y su evaluación, centrada en la evaluación de la usabilidad y la calidad. Se pretende evidenciar que en un indexador se pueden representar y organizar corpus y su documentación y, además, es un sistema cuyo uso es intuitivo y fácil tanto para los autores o creadores del indexador como para los usuarios que necesitan realizar las búsquedas.

La consecución del objetivo general se organiza a través de varios objetivos específicos que se enumeran a continuación.

1. Análisis y recopilación de los requisitos que ha de cumplir el sistema.
2. Formulación del modelo de metadatos descriptivos de los corpus y de las publicaciones relacionadas con estos.
3. Formulación del modelo de catalogación basado en los metadatos descriptivos.
4. Configuración y construcción del repositorio usando el modelo de catalogación y metadatos descriptivos.
5. Implementación de las colecciones de objetos digitales que constituyen el contenido del indexador.
6. Evaluación del sistema mediante indicadores de calidad y un test de usabilidad.
7. Diseño e implementación de mejoras deducidas de los resultados de la evaluación.

¹ Se va a utilizar a lo largo de este documento la palabra *indexador* por mayor precisión semántica, pues el sistema que se ha desarrollado no es un repositorio puro. No obstante, es habitual en la bibliografía que se utilice *repositorio* en un sentido amplio para nombrar a este tipo de sistemas, de forma que en algunas ocasiones también se emplea en este documento como sinónimo.

3 Metodología

El método seguido para lograr los objetivos planteados se basa en el tradicional enfoque de desarrollo de proyectos «en cascada»². Es una metodología de desarrollo secuencial adecuado a este trabajo por la claridad con la que están definidos desde el comienzo del proyecto los requisitos del sistema. Consta de cuatro fases: análisis de requisitos, diseño, implementación y evaluación. A continuación, se especifican estas fases en que se ha dividido el proyecto y las tareas concretas realizadas en cada etapa.

3.1 Fase de análisis

- Análisis funcional: delimitación de las funcionalidades que debe incluir el sistema.
- Revisión de los antecedentes: estudio de los sistemas que pretenden resolver el problema que ocupa a este trabajo, o problemas similares, así como de los métodos que se emplean para ello con el objeto de encontrar la solución adecuada para el indexador ICAE.
- Elaboración de la especificación de requisitos del sistema: explicitación y redacción detallada de los requisitos que ha de cumplir el sistema a desarrollar.

3.2 Fase de diseño

- Formulación del modelo de metadatos descriptivos
 - Modelo entidad-relación
 - Diagrama UML
- Formulación del modelo de catalogación

3.3 Fase de implementación

- Recopilación de datos: recopilación de datos e información entorno a todos los corpus de aprendices de español localizables por internet.
- Configuración del indexador utilizando la aplicación OdA: diseño y definición de la página de portada, definición de los tipos de archivos que puede soportar el sistema, definición de usuarios, definición de página de contacto, configuración del esquema de navegación y las páginas de búsqueda y definición de la página de recomendaciones de uso.
- Implementación de la colección de objetos digitales de tipo *corpus*: inclusión de los metadatos³ y recursos, es decir, de los objetos digitales de la colección.
- Implementación de un prototipo para la colección de objetos digitales de tipo *publicaciones*.

3.4 Fase de pruebas

- Elaboración del plan de pruebas: selección y diseño de las pruebas de evaluación a realizar.
- Evaluación de la calidad: realización de la evaluación basada en criterios de calidad y explicación de resultados.
- Evaluación empírica: realización de la evaluación empírica mediante un test de usabilidad completado por un grupo de usuarios objetivo.

3.5 Fase de mejoras

- Revisión y corrección de errores en el análisis, diseño e implementación.

² Véase Pressman (2010) para más información sobre la metodología en cascada.

³ En OdA, datos del objeto digital, datos del recurso y metadatos del objeto digital.

- Diseño e implementación de mejoras.

3.6 Justificación del proyecto

- Redacción y defensa de la justificación del trabajo.

Las tareas precedentes están representadas cronológicamente en el diagrama que se encuentra en el Anexo I.

El mantenimiento del sistema no está comprendido entre los objetivos de este proyecto. No obstante, se realizará una puesta en marcha del indexador como añadido al resultado del proyecto realizado en este TFM.

3.7 Medios materiales

Los medios materiales necesarios para el desarrollo de este proyecto han sido facilitados por el grupo de investigación UCM-ILSA, que ha cedido un servidor web de aplicaciones, y por la Universidad Complutense de Madrid, que ha concedido al proyecto una beca de colaboración en el Departamento de Lengua Española y Teoría de la Literatura y Literatura Comparada tutorizada por el Dr. José Amenós Pons.

4 Revisión de antecedentes

Los corpus de aprendices son colecciones electrónicas de textos⁴ generados por aprendientes de al menos una lengua meta (LM). Se llama *lengua meta* a «la lengua que constituye el objeto de aprendizaje, sea en un contexto formal de aprendizaje o en uno natural» («Lengua meta», 2008). Es un concepto que engloba los de segunda lengua (L2) y lengua extranjera (LE), aunque en ocasiones se encuentran los tres en la bibliografía como sinónimos⁵.

El corpus de aprendientes es de tipo especializado y se emplea en investigación de adquisición de segundas lenguas⁶, pues facilita una mejor comprensión de la interlengua de los estudiantes. También es un recurso utilizado en docencia para desarrollar material didáctico.

Hay múltiples maneras de organizar la información entorno a estos corpus y todas ellas podrían utilizarse como respuesta al problema que se plantea en este trabajo: listas, tablas, bases de datos, páginas web, repositorios, etc. A continuación, se hace un repaso por los sistemas que se han empleado para dar respuesta a este problema de indexación de corpus de aprendientes de diversas lenguas meta o de otros tipos de corpus.

En el *Learner corpora around the world*, se ha utilizado una tabla publicada en una página web de la Universidad Católica de Lovaina a la que se puede acceder mediante la URL <https://uclouvain.be/en/research-institutes/ilc/cecl/learner-corpora-around-the-world.html>. Esta tabla recoge 169 corpus con diferentes lenguas meta, de ellas, 19 tienen entre sus lenguas meta la española.

Como se ve en la siguiente ilustración, la tabla cuenta en la primera fila con una serie de campos correspondientes a los metadatos que describen los corpus. Estos son: *título del corpus* (enlazando al propio corpus cuando es posible), *lengua meta*, *lengua materna*, *medio* (escrito,

⁴ La palabra *texto* se utiliza aquí en un sentido amplio que abarca desde textos escritos hasta registros de voz o vídeo.

⁵ En este documento se hará distinción entre lengua extranjera y segunda lengua y se empleará *lengua meta* como término general.

⁶ En este trabajo se emplea *adquisición de segundas lenguas* para referirse tanto a procesos de adquisición como de aprendizaje. Para más detalle se puede consultar «Adquisición de segundas lenguas» (2008).

oral...), tipo de texto o de tarea, nivel de dominio de la lengua, tamaño en palabras, director del proyecto y contacto cuando es posible y disponibilidad.

Learner corpora

Search:

Corpus	Target language	First language	Medium	Text type/ task type	Proficiency level	Size in words	Project director	Availability
A Learners' Corpus of Reading Texts	English	French	spoken	Unprepared reading of English texts. The texts are short abstracts of fiction or made-up dialogues.	University students		Sophie Herment Valérie Kerfelec Laetitia Leonarduzzi Gabor Turcsan	Freely available
Analyzing Discourse Strategies: A Computer Learner Corpus	German	English (mainly American English)	written	Threaded Discussion Chat Essays Longitudinal data	From beginner to intermediate	Under development	Christina Frei Edward Nixon University of Pennsylvania, USA	
Aprescripv ("Aprendera Escribiren Lovaina")	Spanish	Dutch	written	Written assignments and tests; several text types (letters, expository, descriptive, argumentative, narrative)	A1 to C1	c. 1 m.	Kris Buysse KU Leuven, Belgium	Restricted online access
Arab Learner English Corpus	English	Arabic	written	Essays written by freshman students as part	University students (second	Analysis 184749 Narrative 67527	imahfouz@auk.edu.kw>Dr. Inas Khatib> American University of	https://dspace.auk.edu.kw/handle/11675/1797

Ilustración 1: tabla de datos Learner corpora around the world

Además de la representación de cada corpus en forma de fila de la tabla con los valores de los metadatos citados para cada corpus, el sistema permite dos funcionalidades:

1. Organización de cada columna (por tanto, cada metadato) por orden alfabético directo o inverso de forma que todas las filas se adapten al orden que se le ha dado a una columna, con el objeto de poder localizar más fácilmente los valores concretos que se buscan.
2. Búsqueda por concordancia con el valor deseado, devolviendo las coincidencias de lo que se haya escrito en la caja de búsqueda. Se permite el uso de los operadores lógicos OR y AND en las búsquedas.

Esta forma de centralizar la información tiene algunas limitaciones. En primer lugar, es difícil la exploración de los recursos para alguien que no sabe lo que busca o que no domina el área de conocimiento. En segundo lugar, no es posible hacer bien búsquedas usando varios criterios (por ejemplo, buscar corpus orales que tengan como lengua meta el inglés). De hecho, ni siquiera se puede buscar exclusivamente por L1 o por lengua meta, pues debido al funcionamiento de la búsqueda por concordancia el sistema devuelve cualquier corpus entre cuyos valores de metadatos esté la lengua deseada. Es decir, que si se quieren buscar corpus cuyos informantes tengan el inglés como L1 lo que se obtiene como resultado son esos recursos más todos aquellos en los que el inglés sea lengua meta. Además, el número de metadatos elegidos para describir los recursos es bastante reducido y no permite por tanto hacer consultas más precisas.

En cuanto a las búsquedas con operadores lógicos, los resultados que dan no son los correctos. Por ejemplo, si se prueba a buscar «spoken OR written» lo que se obtiene es una lista de corpus cuyo valor en el campo *medium* es «spoken and written», la misma que se logra si se busca «spoken AND written». Además, no está especificado en ninguna sección de la web que se puedan usar estos operadores ni hay unas instrucciones para hacerlo. Se trata de un método complejo de utilizar si no se está familiarizado con él (conocimiento que no es necesariamente esperable de los investigadores o profesionales de la enseñanza de lenguas que usen este sistema).

A parte de la mencionada tabla, se han encontrado otras soluciones en internet que acometen un problema cercano al de este proyecto. A continuación, se hace una breve descripción de cada una apuntando aspectos interesantes que aportan o carencias que tienen:

- https://www.leuphana.de/fileadmin/user_upload/PERSONALPAGES/_abcd/barron_anne/files/Hyperlink - Freely available corpora-updated.pdf

Se trata de varias listas en formato PDF que albergan nombres, enlaces y una breve descripción redactada de multitud de corpus con distintos propósitos. Entre ellas hay una enumeración de cinco corpus de aprendices de varias lenguas y un enlace a la tabla de la UCLovain.

Esta lista solo permite la búsqueda habitual por coincidencias con todo el texto de los documentos PDF. No obstante, el número de corpus que recoge es tan reducido (cinco) que no se hacen necesarias otras funciones de búsqueda o navegación más complejas. Las carencias más evidentes de este sistema son, por lo tanto, el limitado número de corpus de aprendientes que lista, la falta de información sobre ellos y un sistema de búsquedas o navegación que permita la recuperación de los recursos a través de sus metadatos.

- http://martinweisser.org/corpora_site/CBLLinks.html

La página *Corpus-based Linguistics Links* fue creada en 2001 por David Lee y mantenida por Martin Weisser a partir de 2016. Es un sistema que indexa distintos corpus y los organiza mediante categorías basadas, en inicio, en la propuesta de Kennedy (1998):

- corpus preelectrónicos,
- primera generación de corpus
- y segunda generación de (mega)corpus.

Esta clasificación ha sido ampliada por los autores a lo largo del tiempo y actualmente recoge, además de esas tres categorías, *corpus históricos y diacrónicos*, *corpus recientemente escritos*, *corpus orales*, *corpus conversacionales*, *corpus especializados*, *corpus de adquisición de L1 y L2*, *corpus analizados y anotados sintácticamente*, *corpus de inglés online*, *páginas web como corpus*, y *corpus no ingleses, paralelos y multilingües*.

A través de este índice se accede a listas que contienen el nombre, la URL y una descripción de las características de cada corpus.

Lo más dificultoso de este sistema se deriva de la redundancia que implica su diseño. Contiene corpus que se engloban en varias categorías pero, en lugar de archivarlos una sola vez y facilitar su recuperación por distintas vías, lo que se hace es multiplicar el mismo ítem para recogerlo bajo tantas categorías como sea necesario. Esto provoca que, para asegurarse de que se han encontrado todos los corpus de hablantes no nativos deseados, haya que recorrer distintas categorías (por ejemplo, *corpus hablados*, *corpus de adquisición de L1 y L2*, *corpus de inglés online* y *corpus de otras lenguas, paralelos y multilingües*).

Por otro lado, se trata de una recopilación dedicada especialmente al inglés⁷, de forma que no recoge prácticamente ningún corpus de aprendientes de español.

- <https://childes.talkbank.org/access/>

CHILDES es un sistema para compartir y estudiar interacciones conversacionales de niños, proporciona un índice que organiza los corpus que contiene la base de datos por grupos de lenguas y tipos de datos. Una de las categorías a las que se puede acceder es

⁷ En palabras del creador, «The English language bias on these pages will, I hope, be forgiven».

la de corpus bilingües. Esta despliega una tabla que contiene los siguientes metadatos de los corpus: *nombre y enlace, lenguas, rango de edad, número de informantes, formato y comentarios*.

De nuevo, el sistema no permite recuperar los ítems mediante búsquedas complejas, sino solo explorar manualmente la tabla o buscar por coincidencias con el texto que aparece.

- <https://corpora.uni-hamburg.de/hzsk/en/repository-search>

El repositorio HZSK da acceso a 52 corpus. Cuenta con un sistema de búsqueda por coincidencias de palabra y otro sistema de filtrado mediante diferentes etiquetas que responden a las categorías de modalidad, tipo de licencia, tipo de corpus, lengua y palabras clave. Entre estas palabras clave se pueden seleccionar etiquetas como *L2 data, successive bilingualism, adult bilingualism, child bilingualism, simultaneous bilingualism, adult L2 acquisition, child L2 acquisition, learner corpus* o *multilingualism*, obteniendo así listas de elementos que dan acceso a las fichas de catalogación y descripción y a los archivos de cada corpus.

Se trata del sistema más completo de los que se han encontrado, pues el método de búsqueda más filtrado por etiquetas resulta bastante útil e intuitivo. No obstante, solo aloja tres corpus de español como lengua meta.

- <http://www.helsinki.fi/varieng/CoRD/corpora/index.html>

CoRD (*Corpus Resource Database*) proporciona acceso a varios recursos en lengua inglesa, entre ellos un corpus de aprendices de inglés. Aporta un formulario que permite filtrar los resultados modificando los valores de varios de los campos que catalogan y describen los corpus. Los metadatos mediante los que se puede acotar la búsqueda son: *nombre, periodo, cantidad de palabras, muestras de textos, corpus oral o escrito, anotación, formato y disponibilidad*. Además, el sistema contiene una ficha con más información por recurso de la que aparece en la tabla de búsquedas. Es similar al anterior en utilidad, pero con menos contenido de corpus de aprendices de lenguas y ninguno en español.

No cabe duda de que los sistemas presentados centralizan el acceso y algunos datos de los corpus de aprendices y, por lo tanto, facilitan su recuperación. Sin embargo, varios de ellos no aportan una información completa, no facilitan la exploración de los recursos, no ofrecen un sistema de búsquedas complejas y la mayoría contienen pocos o ningún corpus en español. En consecuencia, no se puede afirmar que proporcionen una solución completa al problema concreto que se plantea en este trabajo, pero sí son modelos que tener en cuenta. Las ventajas e inconvenientes de las soluciones presentadas han influido, por tanto, en la construcción del Indexador de Corpus de Aprendices.

Con la intención de desarrollar un sistema que mejore estas soluciones para la organización y acceso a los corpus de aprendientes de español se han hecho una serie de elecciones tanto de sistemas de descripción y catalogación de elementos como de tecnologías. Se ha elegido la forma del indexador, se ha utilizado un modelo de facetas jerarquizadas para organizar las colecciones de objetos digitales que forman el sistema y todo ello se ha implementado mediante la aplicación *software* ODA. Seguidamente se explican estas decisiones y se hace un breve repaso de sus fundamentos teóricos.

4.1 Modelos de documentación de corpus y publicaciones

Al no haber encontrado un estándar concreto para la descripción y catalogación de corpus de aprendices, se ha creado un modelo de metadatos expresamente adaptado a la tarea, es decir,

un esquema de metadatos propietario y de propósito específico. La selección de estos metadatos no ha sido arbitraria, sino que se ha logrado como resultado de las fases de análisis y diseño.

La propuesta final se ha alcanzado mediante los requisitos iniciales, un repaso bibliográfico de las propuestas de clasificación de corpus (véase a este respecto Faya, 2015), un recopilatorio de los metadatos aportados habitualmente por los corpus de aprendices⁸ y las necesidades surgidas de la experiencia de implementar el indexador.

El modelo que se presentará en los resultados, en cuanto a las publicaciones relacionadas con los corpus, sigue la propuesta de metadatos para publicaciones académicas de Martínez Arellano y Palma Peña (2017) basada en el estándar Dublin Core.

Para elegir los metadatos que mejor describieran los corpus, se ha intentado además tener en cuenta que la investigación mediante corpus de aprendices de lenguas es interdisciplinaria y reúne lingüística de corpus, adquisición de segundas lenguas, enseñanza, lexicografía, PLN, estudios de traducción, etc. (Granger y Paquot, 2017).

4.2 Indexador

Un indexador es un recurso cuyo objetivo es recopilar, organizar y centralizar en un mismo lugar información dispersa. El indexador, por consiguiente, no tiene como propósito almacenar el contenido de los recursos que indexa sino la información necesaria para encontrarlos en sus lugares de origen. Aunque en la bibliografía sea habitual el uso del término *repositorio* para designar a los indexadores, esta última característica es precisamente la que lo diferencia de un repositorio puro. Un repositorio es «una aplicación web para la gestión y almacenamiento organizado de contenido digital, normalmente en forma de objetos digitales (metadatos + contenido)» (Fernández-Pampillón, 2018). Un repositorio, por ende, aloja el contenido.

Por lo demás, el procedimiento que se ha seguido y las herramientas que se han utilizado en este proyecto para configurar e implementar el indexador son muy similares a lo que se habrían usado para crear un repositorio. Ambos sistemas están además enfocados hacia un mismo objetivo: facilitar al máximo la localización y reutilización de los contenidos o recursos informacionales (Fernández-Pampillón, 2017).

Un ejemplo paradigmático de indexador en el contexto español es el de Hispana, <http://hispana.mcu.es/es/estaticos/contenido.cmd?pagina=estaticos/presentacion>. Se trata de un portal que recolecta información y da acceso a más de siete millones de objetos digitales de más de 200 repositorios, entre ellos las bibliotecas digitales de las comunidades autónomas, los repositorios institucionales de las universidades españolas y los contenidos de la Red Digital de Colecciones de Museos de España.

Hispana dispone de un sistema de búsqueda por *cualquier campo, título, autor, tema, descripción, año o periodo de publicación y colección*. Además, permite búsquedas más complejas mediante el uso de los operadores lógicos *Y*, *O* y *NO*. Entrando en cualquiera de los resultados se llega a una completa ficha del recurso indexado, así como a sus metadatos en

⁸ Una propuesta de estandarización de estos metadatos de corpus de aprendices de LM se puede ver en Granger y Paquot (2017).

diferentes formatos estándares y a las URL que dan acceso al recurso en sí en su repositorio de origen.

En el ámbito europeo hay que destacar Europeana, accesible a través de la URL <https://www.europeana.eu/portal/es>. Es tanto indexador como repositorio, ya que en algunas ocasiones aloja en sus servidores los recursos y en otras ofrece enlaces al exterior del portal. Europeana pone a disposición del usuario casi sesenta millones de objetos en forma de texto, imagen, sonido, vídeo y representaciones en 3D provenientes de museos, bibliotecas y archivos de varios países de Europa.

Europeana posee un sistema de navegación orientado a acercar las colecciones al usuario a través de una clasificación muy clara de los objetos (*texto, imagen, sonido, vídeo y representaciones en 3D*) y un sistema de filtrado de los resultados por *colección, características del documento digital, licencia de uso, país proveedor, idioma, agregador e institución*. Constituye un punto de acceso único a una gran parte del patrimonio europeo y como tal da a los usuarios la posibilidad de descubrir, localizar y recuperar todo tipo de objetos culturales.

Como se ve, los indexadores hacen posible la centralización de una enorme cantidad de objetos digitales alojados en distintos repositorios. Se trata, por lo tanto, de una solución válida al problema general de la dispersión de información y contenido. De la misma forma, la modalidad del indexador puede ser adecuada para centralizar y organizar la información, así como para facilitar la recuperación de los corpus de aprendices de español si son considerados como objetos digitales (recursos y metadatos).

4.3 OdA

Para la implementación del Indexador de Corpus de Aprendices se ha empleado la versión 2.5 de la herramienta OdA por las características y ventajas que se detallan bajo estas líneas.

OdA es una aplicación web de código abierto desarrollada por el grupo de investigación UCM-ILSA⁹ para la creación y mantenimiento de colecciones de objetos digitales (Fernández-Pampillón y Fernández-Valmayor, 2013). Estas colecciones han tomado distintas formas en los proyectos que han usado la herramienta: repositorio, museo virtual, diccionario, etc.

Una de las mayores ventajas de OdA es que permite «crear de forma inductiva colecciones de objetos digitales en dominios altamente especializados» (Fernández-Pampillón, Gayoso, Sarasa y Sierra, s.f.). Esto quiere decir que el esquema de metadatos se puede modificar en cualquier momento, durante y tras la creación de la colección. Además, facilita la creación de objetos digitales de tipos muy variados porque admite muchos formatos de recursos (imágenes, archivos, páginas web, otros objetos digitales...).

En comparación con otros *softwares* de similar propósito como DSpace, es muy llamativa la flexibilidad que ofrece a la hora de crear los esquemas metadatos y los objetos digitales, es decir, en lo concerniente a la creación de la colección. Es de una gran utilidad la posibilidad de modificar la estructura que define la colección (los metadatos) a medida que se va construyendo o incluso ya construida, porque surgen nuevas necesidades o se ven aspectos que no se habían apreciado en un principio. Permite, además, que un repositorio siempre pueda crecer, cambiar y reflejar perspectivas cada vez más complejas y abstractas de un dominio (Fernández-Pampillón et al., s.f.).

⁹ ...junto con investigadores del Proyecto Arqueológico El Caño (Panamá) y con la empresa Varadero Software Factory.

En esta primera fase de ICAE se ha plasmado la realidad más concreta de los dos tipos de objetos que se han incluido: corpus y publicaciones sobre los corpus. Pero el sistema OdA permite continuar incluyendo en un futuro nuevas perspectivas de estos objetos, probablemente más complejas y abstractas.

Finalmente, OdA ofrece la ventaja de ser una aplicación consolidada que ha demostrado su éxito y utilidad como herramienta de creación de repositorios en diversos dominios. Algunos de los proyectos llevados a cabo con OdA son:

- <http://oda-fec.org/nata-english>
Repositorio Arqueológico del Proyecto El Caño (Panamá),
- http://repositorios.fdi.ucm.es/ciberia_oda
Biblioteca Digital Ciberia,
- <http://museofis.ucm.es/museo/>
Museo de Instrumentos Científicos de la Facultad de Ciencias Físicas (UCM),
- <http://repositorios.fdi.ucm.es/DiccionarioDidacticoLatin>
Diccionario Didáctico Digital de Latín

4.3.1 Colecciones de objetos digitales

Como es habitual en estos sistemas, el contenido (en sentido amplio) se almacena en un repositorio OdA en forma de objetos digitales. Un objeto digital, como se ha dicho, está compuesto por contenido informacional y metadatos (Fernández-Pampillón et al., s.f.). En OdA, el contenido informacional es el recurso del objeto digital y los metadatos se dividen en datos, datos del recurso y metadatos.

Los objetos digitales se organizan en colecciones de objetos digitales. Lo que hace que varios pertenezcan a la misma colección es que comparten un mismo esquema de metadatos. ICAE alberga dos colecciones: corpus y publicaciones asociadas a esos corpus. Cada una se define mediante una serie de metadatos que se presentarán en el apartado de resultados. La manera de relacionar diferentes colecciones de objetos digitales son los recursos asociados a un objeto, pues un objeto digital puede constituirse en recurso de otro.

En el caso de la colección de corpus en ICAE, el recurso suele ser la URL de acceso a los corpus y, en algunos casos, un archivo que contiene el corpus. En cuanto al prototipo de la colección de publicaciones (aún en construcción), se prevén dos recursos posibles: el archivo que contiene la publicación y una URL (u otra forma de acceso) al contenido alojado en el exterior del indexador. Además, los corpus tendrán como recurso adicional el objeto digital de la publicación basada en ellos o en su uso. Como contraparte, las publicaciones también contendrán en forma de recurso el objeto digital correspondiente al corpus del que parten.

Para organizar los metadatos de que se compone un objeto digital y catalogar estos elementos, se ha empleado un modelo de facetas jerarquizadas que se explica a continuación.

4.3.2 Facetas jerarquizadas

OdA emplea un modelo de facetas jerarquizadas y es, por lo tanto, del que se ha servido ICAE. Las *facetas* son las categorías disjuntas que se utilizan para describir la información de los elementos de una colección y cada una tiene asignado un nombre (por ejemplo, *método de cocción* en el caso de una colección de recetas). Las facetas pueden ser planas o jerárquicas. Además, tienen un conjunto de *etiquetas* asociadas (Hearst, 2006b), es decir, de nombres usados dentro de cada faceta (Chandler, Hearst y Smalley, 2006). Se podría decir que las facetas

son distintos aspectos que definen a un elemento, distintas perspectivas desde las que se puede observar un objeto.

Una faceta jerarquizada es una faceta que puede tener subfacetas y etiquetas asociadas dispuestas según un modo jerárquico. Una faceta jerárquica en el ejemplo de una colección de recetas sería *ingredientes > lácteos, vegetales y frutas*. Dentro de *lácteos* habría un conjunto de etiquetas entre las que estarían, por ejemplo, *queso, yogur y leche*.

La estructura de categorías facetadas es «intermedia en complejidad y muy rica en flexibilidad» (Hearst, 2006a) respecto a otros sistemas de organización de la información como son las taxonomías, tesauros u ontologías¹⁰. Esta flexibilidad que ofrece facilita la construcción y la navegación por las colecciones. Como ya se ha apuntado en el apartado anterior, esto es una ventaja junto con la falta de límites para seguir añadiendo facetas, subfacetas y elementos que caracteriza a OdA.

Otra ventaja que tiene este tipo de clasificación es permitir que haya múltiples maneras de recuperar cada ítem. Por ejemplo, si entre un grupo de recetas se quiere encontrar una tarta con fresas sin horno solo habrá que buscar mediante estas etiquetas (*tarta, sin horno*) para hallar resultados como una receta de tarta de fresa hecha con queso y gelatina. Esto da al usuario mucha libertad para explorar y buscar en una colección y le permite navegar a través de las dimensiones conceptuales que describen los ítems (Hearst, Li, Swearingen y Yee, 2003).

Las facetas representadas en ICAE se han obtenido del modelo de metadatos diseñado para el sistema. Se pueden dividir en cuatro para los corpus: *datos sobre el corpus, datos sobre los informantes, datos sobre las lenguas habladas por estos informantes y datos sobre los creadores*. En el caso de las publicaciones solo hay una faceta con datos sobre la publicación por el momento. Estas facetas contienen jerarquías formadas por subcategorías, y cada una de estas subcategorías puede desplegar otra jerarquía o contener valores o etiquetas que se asignan a cada objeto digital (a cada corpus o publicación).

5 Resultados del proyecto y discusión

En este apartado se exponen y discuten los resultados de cada fase del proyecto, relacionándolos con los objetivos específicos correspondientes a cada etapa.

5.1 Resultados del análisis

5.1.1 Especificación de requisitos y revisión de antecedentes

Una parte de los resultados de la fase de análisis se ha plasmado en la especificación de requisitos del sistema que se adjunta a este documento en el Anexo II. La otra parte, correspondiente a la revisión de antecedentes, se ha reflejado en la sección del mismo nombre de este documento.

Mediante la especificación de requisitos se establecen las funcionalidades con las que ha de contar el sistema. En este caso fija también las características (los metadatos) de los corpus que se han de recoger y las búsquedas que se deben poder hacer.

¹⁰ Para más información sobre sistemas de organización de la información se puede consultar Fernández-Pampillón (2017).

Por otro lado, para poder saber qué sistema desarrollar como solución al problema planteado es necesario investigar previamente los antecedentes. De los resultados de esta investigación ya se ha hablado en este escrito.

5.1.2 Discusión

La especificación de requisitos ha sido fruto del análisis del problema respecto a la dispersión de información sobre corpus de aprendices de español como lengua meta planteado en el marco de la colaboración con el Departamento de Lengua Española y Teoría de la Literatura y Literatura Comparada.

Los requerimientos establecidos en esta fase tienen relación, como se ha dicho, con los metadatos que debe recoger ICAE y las acciones que se han de poder realizar, las funciones. Entre estas funciones, son especialmente importantes el sistema de búsquedas y navegación que se debe ofrecer al usuario.

Con los resultados mostrados se alcanza el objetivo 1, el análisis del sistema y recopilación de los requisitos que han de cumplir. No obstante, los sistemas de búsqueda y navegación adolecen de ciertas dificultades no previstas durante el análisis y que se explicarán más adelante.

5.2 Resultados del diseño

5.2.1 Modelo de metadatos

La descripción de los objetos (corpus o publicaciones) indexados se ha hecho mediante un esquema de metadatos propietario y de propósito específico. Esto es, una serie de metadatos que no pertenecen a ningún estándar, que han sido creados *ad hoc* para la colección y que están vinculados a un dominio temático particular. Este modelo se presenta a continuación¹¹:

Metadatos comunes		
Nombre	Descripción	Valores posibles
Título	Nombre otorgado al recurso.	<i>Corpus de Conversaciones en Español como Lengua Extranjera, El corpus de aprendices Aprescrilov y su utilidad para la didáctica de ELE en la Bélgica multilingüe</i>
Tipo de objeto	Clase de objeto digital que determina la colección a la que pertenece.	Corpus, publicación

Tabla 1: metadatos comunes a todas las colecciones

¹¹ Para el diseño del esquema se hicieron primero un modelo conceptual entidad-relación y un diagrama UML que se pueden ver en los anexos II y III.

DESCRIPCIÓN DE CORPUS		
Nombre	Descripción	Valores posibles
Tipo de corpus	Clase de corpus según el origen de los textos.	Oral, escrito, oral y escrito, CMO (comunicación mediada por ordenador)
Formato	Formato de los datos que forman el corpus.	Texto, audio, transcripción, vídeo, audio y transcripción, vídeo y transcripción, audio, vídeo y transcripción
Tamaño	Volumen de elementos que forman el corpus.	Número de palabras, número de frases, número de textos, número de conversaciones, número de grabaciones, horas de audio o vídeo, minutos de audio o vídeo
Objetivos	Finalidad con la que se ha creado el corpus.	(Descripción del objetivo para el que se ha creado)
Forma de acceso	Requerimientos impuestos para acceder al corpus.	Libre acceso, libre acceso para investigadores, restringido, restringido (contacto con el autor), disponible en CD-ROM, no disponible, no terminado
Combinación de lenguas meta	Lenguas de los textos que forman el corpus de aprendientes.	Monolingüe, bilingüe, multilingüe Español, español / francés, árabe / español / inglés
Tipo de tarea	Tipo de actividad que se ha llevado a cabo para generar los textos que forman el corpus.	Conversación, conversación y lectura, ensayos, entrevista; narración; narración, entrevista, descripción; palabra; redacción; trabajo académico; varios

Tipo de texto	Clase de texto según su estructura.	Argumentativo, dialogal, narrativo, descriptivo, expositivo
Fuente	Origen de los textos que forman el corpus.	Ejercicio de clase, ejercicio de clase y prueba escrita, ejercicio <i>online</i> , grabación de audio, grabación de audio y vídeo, grabación de vídeo, prueba escrita, prueba oral y escrita, varias
Fecha de obtención	Fecha o periodo temporal en que se han producido los textos que forman el corpus.	1999-2000, 2004, en curso
Fecha de publicación	Fecha o año de publicación del corpus.	11/03/2009, 2009, no publicado
Anotación	Tipo de anotación del corpus.	Anotado, errores anotados, sin anotar
Número de informantes	Número de personas que han generado los textos del corpus.	(número entero)

Tabla 2: metadatos de corpus

Descripción de informantes		
Nombre	Descripción	Valores posibles
Edad	Edad o rango de edad de los informantes en el momento de la obtención del corpus.	Niños, adultos (Número entero, rango de números)
Nivel de español	Nivel de lengua de los informantes en el momento de la obtención del corpus.	Bajo, bajo-medio, medio, medio-alto, alto, otros
Nivel de estudios	Grado de estudios alcanzado por los informantes en el momento de la obtención del corpus.	Ciclo formativo de grado superior, diverso, secundaria, secundaria y superior, superiores
Género	Grupo al que pertenecen los informantes según su sexo.	Hombre, mujer, ambos
Profesión	Dedicación profesional o de otra índole de los informantes en el momento de la obtención del corpus.	Estudiante

Nivel sociocultural	Estatus de los informantes en la sociedad según sus conocimientos culturales.	Medio-alto
----------------------------	---	------------

Tabla 3: metadatos de informantes del corpus

Descripción de lenguas		
Nombre	Descripción	Valores posibles
L1	Lengua materna de los informantes.	Inglés, Alemán / español / francés / italiano / portugués
Bilingüismo	Lenguas en las que son bilingües los informantes.	Árabe, Español / inglés
L2	Segunda lengua o lengua extranjera principal de los informantes.	Español, Español / inglés / francés
Otras lenguas	Otras lenguas que hablan los informantes.	Catalán, Alemán / árabe / catalán / francés / inglés / portugués / ruso

Tabla 4: metadatos de lenguas de los informantes

Descripción de creadores		
Nombre	Descripción	Valores posibles
Autoría y créditos	Nombre del autor o dueño del corpus.	(Nombres de personas, equipos, instituciones o cualquier otra entidad)
Editor	Nombre del editor del corpus	(Nombres de personas, equipos, instituciones, revistas o cualquier otra entidad)
Contacto	Dirección electrónica de contacto con el autor o dueño del corpus.	(Dirección de correo electrónico)
Rol del investigador	Papel jugado por el investigador en relación con los informantes durante la obtención del corpus.	Presente, presente con poca intervención, no presente, observador no participante

Tabla 5: metadatos de creadores

Descripción de publicaciones asociadas		
Nombre	Descripción	Valores posibles
Tipo de publicación	Clase de publicación académica según su estructura, extensión y formato.	Artículo, libro, capítulo, ponencia

Autor/es	Nombre del autor o autores de la publicación.	(Nombre de persona o entidad)
Traductor/es	Nombre del traductor o traductores de la publicación.	(Nombre de persona o entidad)
Localización	Publicación mayor en que se encuentra la publicación.	(Nombre de revista o libro)
Fecha de publicación	Fecha en que se ha publicado el documento.	(Fecha)
Resumen	Exposición resumida de la publicación.	(Texto)
Idioma	Lengua en que está escrita la publicación.	Español, inglés
Temas	Temas que trata la publicación	ELE, educación, adquisición de segundas lenguas

Tabla 6: metadatos de publicaciones

5.2.2 Modelo de catalogación

El modelo de catalogación está formado por facetas jerarquizadas y se ha construido con base en el esquema de metadatos ya presentado y en vocabularios creados exprofeso para el indexador:

Modelo de Datos	
 TÍTULO	 Número de informantes
 TIPO OBJETO	 Informantes
 CORPUS	 Lenguas
 Tipo de corpus	 L1
 Formato	 Bilingüismo
 Tamaño	 L2
 Objetivos	 Otras lenguas
 Forma de acceso	 Edad
 Combinación de lenguas	 Niños
 Monolingüe	 Adultos
 Bilingüe	 Diversa
 Multilingüe	 Nivel de español
 Tipo de tarea	 Bajo
 Tipo de texto	 Bajo-medio
 Fuente	 Medio
 Creador/es	 Medio-alto
 Autoría y créditos	 Alto
 Editor	 Otros
 Contacto	 Nivel de estudios
 Rol del investigador	 Sexo
 Fechas	 Profesión
 Fecha de obtención	 Nivel socio-cultural
 Fecha de publicación	 Publicaciones asociadas
 Otros rasgos	
 Anotación	

Tabla 7: modelo de catalogación

 PUBLICACIONES
 Tipo de publicación
 Autor/es
 Traductor/es
 Localización
 Fecha de publicación
 Resumen
 Idioma
 Temas
Modelo de MetaDatos
 Creador/a del objeto digital
 Fecha de creación del objeto digital
Modelo de Datos de los Recursos
 Descripción del recurso

Tabla 8: modelo de catalogación

5.2.3 Discusión

La propuesta final de los modelos de metadatos que se presenta en estos resultados cumple con las características de los corpus cuya representación se estableció en la especificación de requisitos (Anexo II) como obligatoria y que se enumeran bajo estas líneas:

De los corpus: título, dirección web¹², fuente de información (especialmente cuando el corpus no está en línea), tipo de corpus, formato, tamaño, objetivos, forma de acceso, combinación de lenguas, tipo de tarea, tipo de texto, fuente, fecha de obtención, fecha de publicación, otros rasgos (anotación, por ejemplo), número de informantes, publicaciones relacionadas; del creador o creadores: autoría y créditos, editor, contacto, rol del investigador; de los informantes: L1, bilingüismo, L2, conocimiento de otras lenguas, edad, nivel de español, nivel de estudios, género, profesión, nivel sociocultural.

Incluso se ha mejorado y enriquecido esta lista de características representando las publicaciones como otra colección de objetos digitales con sus propios metadatos y fichas de catalogación.

Además del análisis de requisitos y el estudio de la bibliografía, en el diseño de estos esquemas ha sido clave la flexibilidad de la aplicación OdA para modificar, eliminar y añadir elementos del modelo de metadatos para la construcción de la propuesta final de descripción y catalogación de los objetos digitales.

¹² La dirección web de acceso al corpus o publicación no se incluye como dato o metadato sino como recurso.

No obstante, el empleo de OdA también ha impuesto condiciones sobre la descripción y catalogación de los corpus. El principal condicionante por el que se ha visto afectado ICAE es que la herramienta no permite la implementación de atributos multivaluados. De tal forma que los metadatos *tipo de tarea, tipo de texto, fuente, L1, L2, bilingüismo, otras lenguas, combinación de lenguas y nivel de español* no han podido ser representadas adecuadamente, pues son campos que pueden tener varios valores. Las soluciones encontradas a este problema han sido las siguientes:

- Para *nivel de español*, dividir la faceta en varias categorías «paraguas»: *bajo, medio-bajo, medio, medio-alto, alto, otros*. A cada una de estas categorías se le ha asignado como conjunto de valores un vocabulario controlado en lugar de texto con la intención de facilitar la búsqueda, pues los vocabularios controlados en OdA están formados por términos concretos que en la búsqueda aparecen en un desplegable de opciones. Tener que elegir entre varias alternativas en lugar de escribir asegura que la búsqueda dé resultados.

- Para las subfacetas que forman la jerarquía de *lenguas*, la de *combinación de lenguas* y para *tipo de texto*, se han estandarizado ortográficamente los valores que pueden tener. La estandarización consiste en que las enumeraciones de los valores (términos) están ordenadas por estricto orden alfabético; los términos están separados por un espacio, una barra (/) y otro espacio (*alemán / francés*) y respetan las reglas de acentuación del español.

Esta estandarización, sin embargo, sigue sin permitir la búsqueda cómoda y directa de corpus con varias lenguas meta porque no se puede prever en qué orden aparecerán en la enumeración debido a que este varía en función de las lenguas correspondientes a cada corpus. La solución ideal, que una faceta o metadato tenga varios valores, no es posible en OdA. No obstante, es una solución de compromiso que funciona aceptablemente.

Otras opciones que se han barajado como crear un vocabulario controlado con todas las combinaciones posibles de valores o desplegar las L1 en distintos campos (L1_1, L1_2...) presentaban muchos problemas. Así, en el caso de la segunda alternativa, habría sido muy confuso que, en un corpus cuyos informantes tuvieran como L1 el alemán y el francés, alemán apareciera como valor del campo L1_1 y francés como valor del campo L1_2 y en un corpus con las L1 albanés, alemán y francés, albanés fuese L1_1, alemán, L1_2 y francés, L1_3. Esto generaría redundancia en el sistema y confusión en los usuarios, de forma que se considera que no es una solución apropiada al problema de las subfacetas multivaluadas.

- Para las facetas *tipo de tarea y fuente*, se ha decidido utilizar un vocabulario controlado con las distintas combinaciones posibles, pues al haber pocas combinaciones los valores son más claros que en los casos anteriores.

Otro requisito de OdA que explica parte de la organización que aparece en el modelo de catalogación es el de dividir los metadatos en tres tipos: (i) los que conforman el esquema de datos, es decir, los metadatos que describen el recurso que se va a encontrar en el objeto digital

—el corpus o la publicación, en este caso—; (ii) los correspondientes al esquema de metadatos, referido a los metadatos del propio objeto digital, en este caso el autor del objeto y la fecha en que se creó y (iii) el esquema de datos del recurso (en este caso, la descripción del tipo de recurso concreto que se ha incluido en el objeto digital).

Asimismo, OdA no admite la creación de colecciones separadas en un mismo repositorio con esquemas de datos/metadatos diferenciados para distintas colecciones. Para solventar esto, lo que se ha decidido hacer es dividir el macrosquema en dos apartados (corpus y publicaciones) que contengan cada uno sus metadatos correspondientes. Así, cuando se quiere incluir un objeto digital de tipo *corpus*, se rellenan solo los campos correspondientes a este apartado, y lo mismo con una publicación.

Lo que se ha hecho para relacionar una colección con otra ha sido añadir al corpus en cuestión un recurso en forma de objeto digital que sea *publicación asociada* y otro que sea *corpus base* en el caso del esquema de *publicaciones*¹³.

A pesar de las limitaciones comentadas, la comparación de los modelos de metadatos y de facetas propuestos en este proyecto con el modelo de la tabla de la UCLovain, comentada en la revisión de antecedentes como sistema que intenta resolver el mismo problema que ICAE, es claramente favorable al indexador. En ICAE se proponen 40 metadatos organizados mediante facetas jerarquizadas frente a los 9 campos que ofrece la UCLovain. Como el modelo de ICAE es mucho más completo, permite más precisión en las búsquedas u obtener más información en la navegación sin necesidad de entrar en las páginas de los corpus o de buscar otras fuentes de información.

Por otro lado, siempre cabe la posibilidad de mejorar el modelo añadiendo más campos que cataloguen y describan los corpus y es algo que se podrá hacer en próximas etapas del proyecto gracias a la ya mencionada flexibilidad de OdA. Se podrá, sobre todo, ampliar las facetas que describen estos recursos para representar otros aspectos como los tipos de errores que contienen los corpus o los recorridos didácticos en los que se podrían utilizar. Sin embargo, estas ampliaciones exceden el alcance del presente trabajo, que abarca, como se puede ver en sus objetivos, la descripción de los objetos reales que constituyen los corpus y la implementación de este sistema, así como una propuesta de descripción de las publicaciones asociadas a los corpus y un prototipo de implementación.

Dicho lo anterior, se concluye que el modelo de metadatos y las fichas de catalogación generadas a partir de él están compuestos de campos adecuados y tienen una extensión también coherente con la información que suelen ofrecer los creadores de este tipo de corpus sobre el mismo. Se puede afirmar, además, que supone mejoras y aporta una mayor riqueza respecto del modelo que utiliza la tabla de UCLovain.

¹³ Esto se puede comprobar en los objetos 327 (corpus) y 347 (publicación).

5.3 Resultado de la implementación

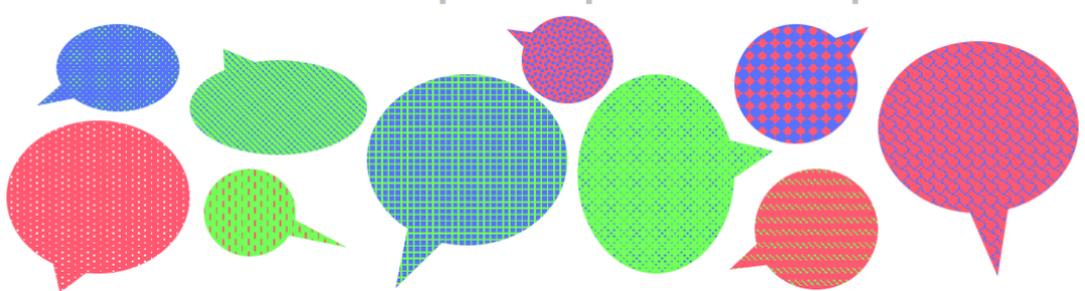
5.3.1 El Indexador ICAE

Se puede acceder al resultado de la implementación del indexador (es decir, al propio indexador) a través de la siguiente URL: http://repositorios.fdi.ucm.es/corpus_aprendices_español.

Se encuentra en primer lugar la presentación (Ilustración 2) compuesta por una ilustración y un texto con información sobre la aplicación y la manera de utilizarla.

> Presentación del sitio web

Indexador de Corpus de Aprendices de Español



El Indexador de Corpus de Aprendices de Español es una aplicación web que recopila y centraliza información sobre los corpus de aprendices de español como lengua extranjera y como segunda lengua.

El objetivo de este recurso es el de facilitar la tarea investigadora y docente a aquellas personas que necesiten consultar este tipo de corpus. Se trata de presentar a los profesionales de ELE un solo recurso que permita acceder a toda la información sobre los corpus existentes. De manera que no pasen desapercibidos corpus que pueden resultar de interés para la investigación o la labor docente a la vez que se simplifica la tarea de búsqueda.

Actualmente la aplicación contiene una colección de corpus en forma de objetos digitales. Estos objetos están formados por una serie de datos, metadatos y recursos asociados. Los datos y metadatos describen, en este caso, cada corpus permitiendo su localización. Los recursos son los enlaces, documentos, imágenes, etc. asociados a cada objeto digital. En este indexador suelen ser enlaces a otras páginas web y, en alguna ocasión, archivos de texto.

Hay tres formas de recuperar los corpus: a través de la navegación, explorando las categorías que los describen; mediante la búsqueda, rellenando los campos de un formulario para hacer búsquedas simples o complejas, y accediendo de forma secuencial a los objetos mediante la pestaña *colecciones*.

El indexador recoge tanto corpus ya publicados como otros que están aún en construcción. También incluye en algún caso corpus de los que hay información en internet pero no mucha claridad respecto al estado en el que están. El motivo de recoger estos trabajos a medias es facilitar a quien lo desee el contacto de los creadores de manera que se pueda solicitar más información.

Ilustración 2: presentación de ICAE

ICAE consta de varias formas de acceso a los contenidos que alberga, a saber:

- Navegación (ilustraciones 3 y 4): se pueden explorar los contenidos a través de un menú de navegación creado a partir de los metadatos seleccionados para este fin. En este caso son:
 - Título, tipo de objeto, corpus, publicaciones.
 - Corpus > tipo de corpus, formato, combinación de lenguas, tipo de tarea, tipo de texto, fuente, informantes.



Ilustración 3: navegación clasificada de la colección de corpus

- Publicaciones > tipo de publicación, temas.



Ilustración 4: navegación clasificada de la colección de publicaciones

- Buscador (ilustración 5): se trata de una presentación de los objetos digitales de manera secuencial que permite además buscar por coincidencias con el título de los recursos.

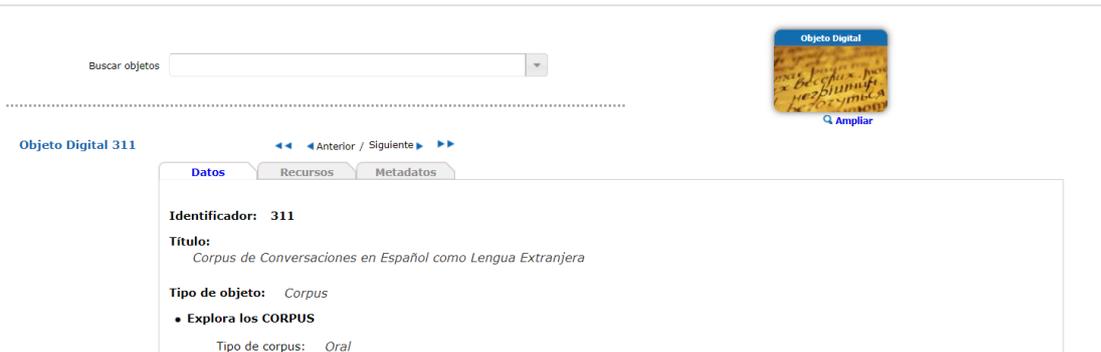


Ilustración 5: apartado Colecciones

- **Búsqueda:** es un sistema de búsqueda basado en un formulario cuyos campos se definen mediante el esquema de metadatos.

[Buscar](#)

Modelo de Datos

Título	<input type="text"/>
Tipo de objeto	= ▼ -- Tipo de objeto -- ▼
Explora los CORPUS	
Tipo de corpus	= ▼ -- Tipo de corpus -- ▼
Formato	= ▼ -- Formato -- ▼
Tamaño	<input type="text"/>
Objetivos	<input type="text"/>
Forma de acceso	= ▼ -- Forma de acceso -- ▼
Combinación de lenguas meta	
Monolingüe	<input type="text"/>
Bilingüe	<input type="text"/>
Multilingüe	<input type="text"/>
Tipo de tarea	= ▼ -- Tipo de tarea -- ▼
Tipo de texto	<input type="text"/>
Fuente	= ▼ -- Fuente -- ▼
Creador/es	
Autoría y créditos	<input type="text"/>
Editor	<input type="text"/>
Contacto	<input type="text"/>
Rol del investigador	<input type="text"/>
Fechas	
Fecha de obtención	<input type="text"/>
Fecha de publicación	<input type="text"/>
Otros rasgos	
Anotación	= ▼ -- Anotación -- ▼
Número de informantes	= ▼ <input type="text"/>

Ilustración 6: muestra parcial del apartado Búsqueda

Los contenidos a los que se accede por las vías explicadas son los objetos digitales, 37 en total (36 corpus y un prototipo de publicación). Los objetos digitales en ICAE tienen la siguiente forma:

Buscar objetos:

Objeto Digital  Ampliar

Objeto Digital 310 ◀ Anterior / Siguiente ▶▶

Datos | Recursos | Metadatos

Identificador: 310

TÍTULO:
Corpus de textos escritos por alumnos tunecinos de enseñanza secundaria estudiantes de E/LE

TIPO OBJETO: Corpus

• **CORPUS**

Tipo de corpus: Escrito

Formato: Texto

Tamaño:
4 páginas, 23 redacciones

Objetivos:
Trabajo de investigación Análisis de la influencia del francés como L2 en el uso y aprendizaje de las preposiciones del español por parte de alumnos tunecinos de lengua materna árabe

Forma de acceso: Libre acceso

• Combinación de lenguas

Monolingüe:
Español

Tipo de tarea: Redacción

Fuente: Ejercicio de clase

• Creador/es

Autoría y créditos:
Imen Boudali, Universidad de Alcalá

Editor:
LinRed

Contacto:
abdel_imen@hotmail.com

• Fechas

Fecha de obtención:
Antes de 2005

Fecha de publicación:
03/04/2006

• Otros rasgos

Anotación: Sin anotar

Número de informantes: 23

• Informantes

• Lenguas

Bilingüismo:
Francés

L2:
Español

• Edad

Adultos:
15-18

• Nivel de español

Otros:
Segundo y tercero de la escuela secundaria

Nivel de estudios: Secundaria

Sexo: Ambos

Profesión:
Estudiante

Ilustración 7: pestaña de datos de un objeto digital de ICAE

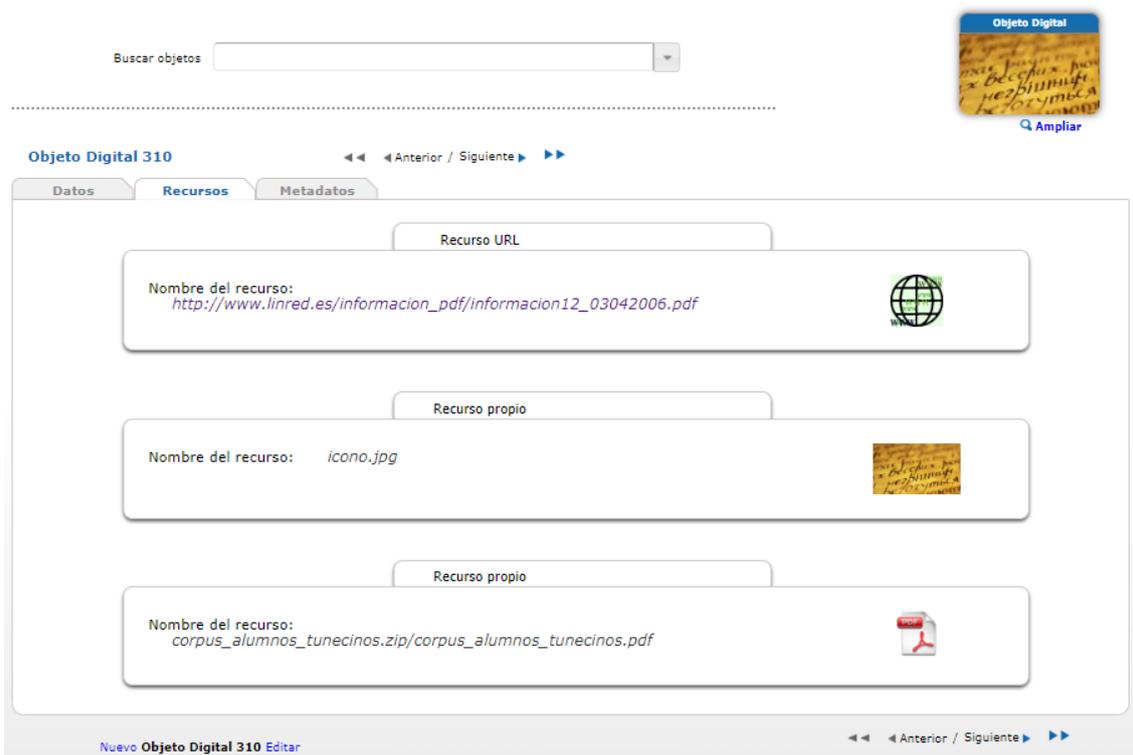


Ilustración 8: pestaña de recursos de un objeto digital de ICAE

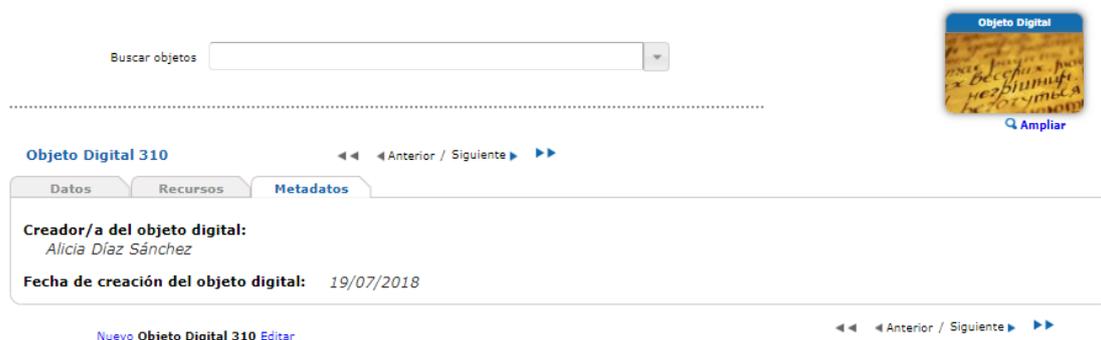


Ilustración 9: pestaña de metadatos de un objeto digital de ICAE

Como se ve en las imágenes, el objeto digital está formado por las siguientes secciones:

- *Datos* (ilustración 7), que recoge los metadatos asignados al corpus o publicación concretos y que sirven para describirlo y catalogarlo;
- *Recursos* (ilustración 8), donde se encuentran la dirección URL que da acceso al corpus o publicación o a la fuente de información de la que se han extraído sus metadatos y, en algunos casos, los documentos que conforman el objeto corpus o publicación. También se pueden hallar descripciones del recurso.
- *Metadatos* (ilustración 9), que alberga los metadatos del objeto digital, es decir, el creador del objeto y la fecha de creación.

Finalmente, hay en el indexador una página de contacto (ilustración 10) donde se encuentran las direcciones de correo electrónico de los administradores, con el objeto de poder ser contactados para incluir otros corpus, para recibir sugerencias o para resolver cualquier tipo de duda que pudieran tener los usuarios. También se incluye un apartado de recomendaciones de

uso (ilustración 11) en el que actualmente se explica el estándar ortográfico que siguen algunos de los datos para que los usuarios puedan tenerlo en cuenta en sus búsquedas.

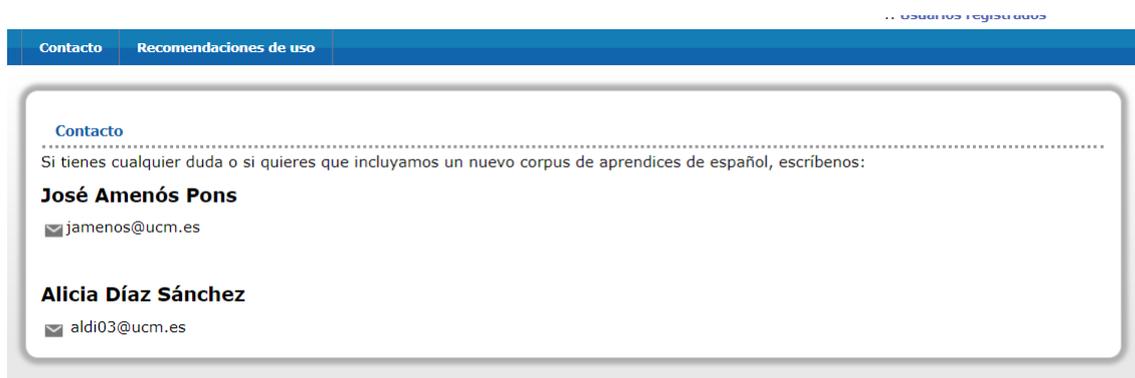


Ilustración 10: página de contacto de ICAE

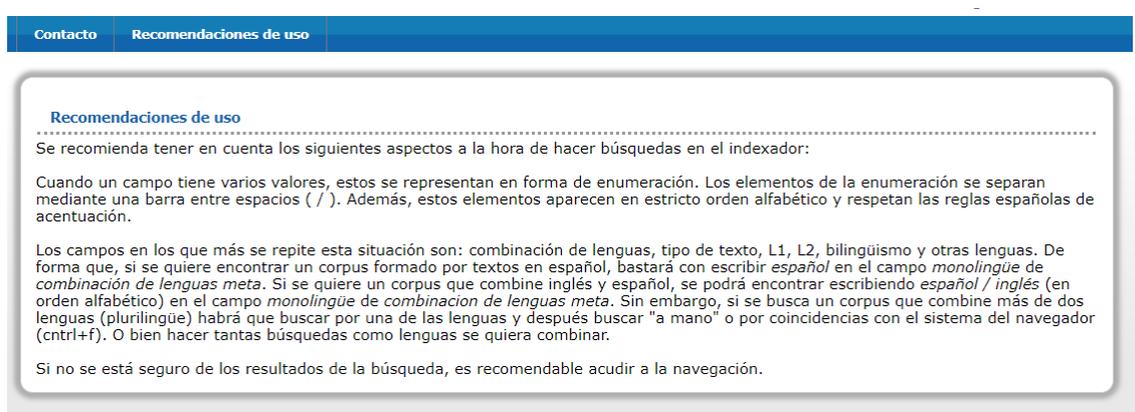


Ilustración 11: página de recomendaciones de uso de ICAE

5.3.2 Discusión

La implementación de este sistema ha consistido, por un lado, en configurar todo el indexador según las posibilidades de OdA: añadir los modelos de datos y metadatos fruto del diseño, generar el sistema de navegación y de búsqueda basados en estos modelos y configurar tanto la página principal de ICAE en el indexador como la página de contacto. Y, por otro lado, en recopilar e incluir en el indexador la información (en forma de valores de los campos pensados para el esquema de metadatos y las fichas de catalogación) y el acceso (en forma de recurso) de los 36 corpus seleccionados. Se ha cumplido así con las tareas propuestas para la fase de implementación y con el objetivo específico número 5: implementación de las colecciones de objetos digitales que constituyen el contenido del indexador.

Los criterios de selección que se han seguido para recopilar los corpus a indexar han sido amplios. Se han admitido corpus tanto de acceso abierto como restringido, corpus terminados y publicados y corpus no terminados o, terminados, pero no publicados. Tan solo se han omitido aquellos en los que hay evidencias claras de que nunca van a culminar ni va a haber acceso alguno a un material aunque sea parcial¹⁴.

¹⁴ Un ejemplo de este tipo de proyectos fantasma es el del CorpusELE del grupo ESPALEX de la Universidad de Salamanca. La única fuente de información el que aparece es la página

Además de corpus de aprendices de español como lengua extranjera, se han incluido también algunos corpus de producciones de personas bilingües por el interés que pueden tener para la investigación en adquisición de segundas lenguas.

La finalidad de seguir unos criterios tan poco estrictos es ofrecer al usuario la mayor cantidad de información posible, junto con los contactos de los creadores de los recursos, de manera que si los corpus no son accesibles al menos se pueda recurrir a sus autores para obtener más datos en caso de que se tenga interés.

El sistema completo que se ha generado y los tipos de búsquedas y la navegación que lo acompañan superan las posibilidades de búsqueda que ofrece la tabla de UCLovain ya explicadas en la revisión de antecedentes. Además, el sistema principal con el que se compara ICAE incluye solo 19 corpus de aprendices de español como LE frente a los 36 que forman actualmente el indexador.

El sistema de navegación por las facetas jerarquizadas permite que un usuario que no tiene claro lo que necesita o que no domina especialmente el área de conocimiento tenga la oportunidad de explorar las colecciones y descubrir recursos que puedan ser de su interés. Por su lado, la búsqueda por formulario posibilita hacer tanto consultas sencillas (rellenando solo un campo de todo el formulario, por ejemplo) como consultas complejas cruzando criterios.

Como se ha explicado en la discusión de los resultados del diseño, se ha tenido que modificar el esquema de datos por la limitación de OdA respecto a la representación de atributos multivaluados. Esto ha supuesto cambios también en la implementación del sistema. Se han debido borrar y volver a incluir los metadatos *tipo de tarea, tipo de texto, fuente, L1, L2, bilingüismo, otras lenguas, combinación de lenguas y nivel de español*. Con la nueva implementación, se ha enriquecido el sistema de búsqueda y navegación.

Sin duda, hay aspectos que se pueden mejorar. Por ejemplo, sería positivo seguir desarrollando el apartado de recomendaciones de uso y apuntar posibles valores que se pueden introducir en algunos campos para hacer búsquedas. En la misma dirección, también se podría generar un videotutorial para aprender a utilizar el indexador. Seguramente tener a mano una información más completa sobre el uso del sistema haría la tarea más sencilla y rápida a los usuarios.

Por último, se deja «en construcción» la implementación de la colección de publicaciones. De esta colección se ha hecho un prototipo con el objeto de ejemplificar la forma de crear dos colecciones diferentes en OdA, de relacionarlas entre sí y de incluir objetos en esta segunda colección. Este prototipo es el objeto digital número 347, y constituye la publicación asociada al objeto 327.

Presentados y justificados los resultados de la implementación, se puede afirmar que se ha cumplido con los requisitos previstos en cuanto a las características de cada corpus que tiene

<http://diarium.usal.es/espalex/recursos/corpusele/> y, según lo transmitido por uno de los miembros del grupo de investigación, éste desapareció hace años y no hay más datos sobre el corpus que los que aparecen en la URL citada.

que recoger ICAE. Incluso se ha ido más allá, reflejando las publicaciones asociadas a los corpus mediante una segunda colección en construcción que añade valor a la colección principal. Asimismo, se ha implementado el sistema de búsquedas y navegación necesario y la aplicación permite hacer las consultas citadas en la especificación de requisitos, así como cruzarlas para refinar las búsquedas.

Con el indexador ya desarrollado, se puede concluir que este sistema supone un avance en la recuperación de la información de corpus de aprendices de español, si bien no hay que perder de vista que se puede mejorar y que tiene el punto débil de no poder representar bien atributos multivaluados. A continuación, se discuten los aspectos que, a la luz de los resultados de las evaluaciones, no funcionan y cuyos problemas habría que solucionar en posteriores etapas del proyecto.

5.4 Resultados de la evaluación

5.4.1 Evaluación deductiva de calidad

La evaluación de la calidad del indexador se ha hecho mediante una serie de indicadores adaptados a las circunstancias concretas de ICAE y basados en indicadores para la evaluación de la calidad de repositorios. Han debido ser ajustados al proyecto porque ICAE no es un repositorio puro, sino un indexador, porque esta evaluación se ha realizado recién nacido el recurso y porque se trata de un indexador temático y de tamaño reducido y no de un indexador institucional¹⁵.

A continuación se presentan los indicadores elegidos, basados en la propuesta de Serrano, Melero y Abadal (2014) y los resultados de la evaluación deductiva:

	INDICADOR DE CALIDAD	ICAE	¿POR QUÉ?
CONTENIDOS	Acceso abierto	SÍ	El indexador está bajo licencia <i>Creative Commons</i> . Gran parte de los contenidos que indexa (a los que hay acceso) también son de acceso abierto.
	Uso de estándares abiertos	NO	La especificidad de los objetos que alberga no permite usar metadatos estandarizados puesto que no existe aún un esquema estándar adecuado.
	Tipo de contenido (patrimonial, académico, institucional...)	SÍ	Indexa un tipo de contenido habitual en los repositorios: académico, material de investigación.
	Política sobre el archivo en el repositorio (documento público)	NO	Mejora por implementar.

¹⁵ En gran parte de los artículos sobre indicadores de calidad de repositorios los repositorios institucionales tienen un mayor protagonismo. Por ejemplo:

- https://recolecta.fecyt.es/sites/default/files/contenido/documentos/GuiaEvaluacionRecolecta_v.ok_0.pdf
- <https://recyt.fecyt.es/index.php/anadoc/article/viewFile/32127/17075>.

MARKETING	Promoción en la institución	SÍ	Están previstas dos sesiones de presentación a alumnos de la Universidad Complutense de Madrid y la promoción del indexador en las páginas de otros proyectos de investigación de la UCM.
	Hipervínculos	NO	Mejora por implementar.
	Otras acciones (promoción externa, blog, informe anual, mecanismos para incentivar el archivo)	NO	Mejora por implementar.

Tabla 9: resultados de la evaluación de calidad

Leyenda: Cumple con el indicador No cumple con el indicador

5.4.2 Evaluación empírica de la usabilidad

La evaluación empírica se ha llevado a cabo mediante un cuestionario creado a propósito para evaluar el indexador y que se puede consultar en el Anexo V. Este cuestionario está formado por una serie de pruebas que consisten en ejemplos de búsquedas que se pide a los usuarios que realicen y por un test de usabilidad más dos afirmaciones añadidas.

Como test de usabilidad se ha utilizado el Sistema de Escalas de Usabilidad creado por John Brooke en 1986. Se ha obtenido de Davin (2017). A él se han añadido dos afirmaciones que no se pueden medir con el resto porque no se rigen por el mismo algoritmo y falsearían los datos obtenidos.

De la primera parte, se presentan los resultados de la evaluación de la facilidad de cada búsqueda. De la segunda, los resultados del test de usabilidad y, por separado, los resultados de las dos últimas afirmaciones. El cuestionario que se ha empleado se puede consultar en el Anexo V y el conjunto total de los datos, en los anexos VI y VII.

Esta evaluación se ha hecho, por un lado, con alumnos de tercero y cuarto del grado de Español y del grado de Lingüística y Lenguas Aplicadas y, por otro, con alumnos del máster en Español como Segunda Lengua. Al grupo de grado en Lingüística se le llamará grupo 1 y al de máster en Español como Segunda Lengua, grupo 2. Los datos resultantes se presentan, por tanto, separados en estos dos grupos.

5.4.2.1 Resultados del cuestionario de usabilidad

El resultado promedio obtenido en el grupo 1 ha sido 47,6 y en el grupo 2, 44,2¹⁶. Aunque en la zona alta del rango, ambos resultados se encuentran en el grado F de la escala de normalización de Sauro y Lewis (2016) que se presenta a continuación.

Grado	Rango de puntuación numérica
A+	84.1-100

¹⁶ No se presentan gráficamente los resultados del cuestionario de usabilidad porque «las puntuaciones independientes no son significativas por sí mismas» (Calvo-Fernández Rodríguez, Ortega Santamaría y Valls Sáez), sino que lo interesante es la puntuación final que se calcula mediante el algoritmo pensado para la escala.

A	80.8–84.0
A-	78.9–80.7
B+	77.2–78.8
B	74.1–77.1
B-	72.6–74.0
C+	71.1–72.5
C	65.0–71.0
C-	62.7–64.9
D	51.7–62.6
F	0–51.6

5.4.2.2 *Resultados del cuestionario sobre la dificultad de las búsquedas propuestas*

A continuación, se especifican las búsquedas que se pidió hacer a los usuarios y, seguido, dos gráficas con los resultados de la encuesta que se les hizo respecto al grado de dificultad de las búsquedas.

Quieres encontrar todos los corpus con la(s) siguiente(s) característica(s):

SITUACIÓN 1

- La lengua materna de todos o de varios informantes es el alemán

SITUACIÓN 2

- La lengua materna de los informantes es el francés
- Recogen muestras de lengua escrita
- Recogen textos narrativos

SITUACIÓN 3

- Son multilingües (es decir, recogen muestras de más de dos lenguas)

SITUACIÓN 4

- Recogen muestras de inglés y español
- Los informantes solo son mujeres

Las respuestas se han medido mediante una escala de Likert con cinco grados: *muy fácil, fácil, ni fácil ni difícil, difícil y muy difícil*. En el eje vertical de la gráfica aparecen estas cinco categorías y en el horizontal el número de veces que se repite cada opción en las respuestas de los usuarios. Se representan en la misma gráfica las cuatro búsquedas identificadas con barras de distintos colores. De esta manera se puede ver el número de usuarios a los que les ha resultado fácil la búsqueda 1, a cuántos les ha resultado difícil la búsqueda 3, etc.

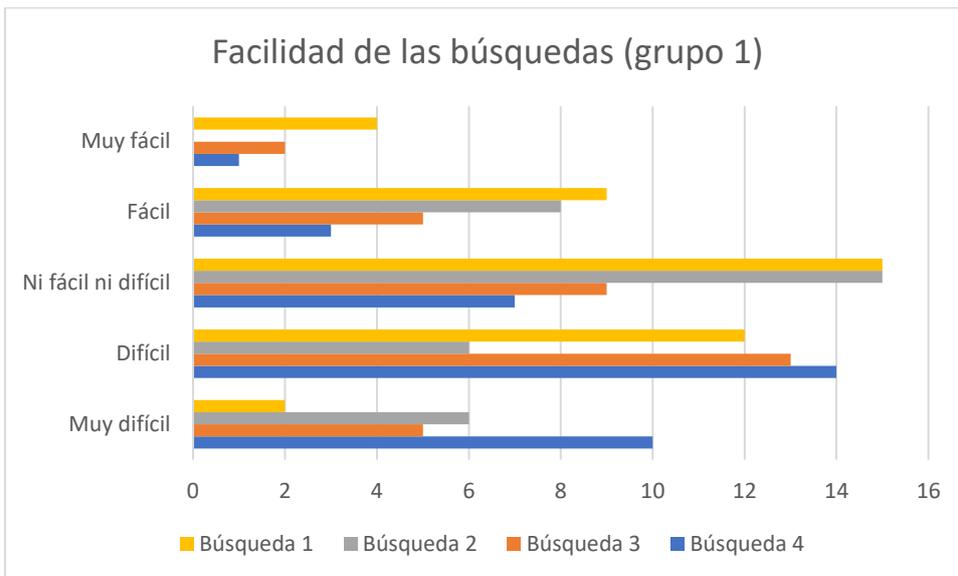


Gráfico 1: resultados de la encuesta sobre la facilidad de las búsquedas propuestas (grupo 1)

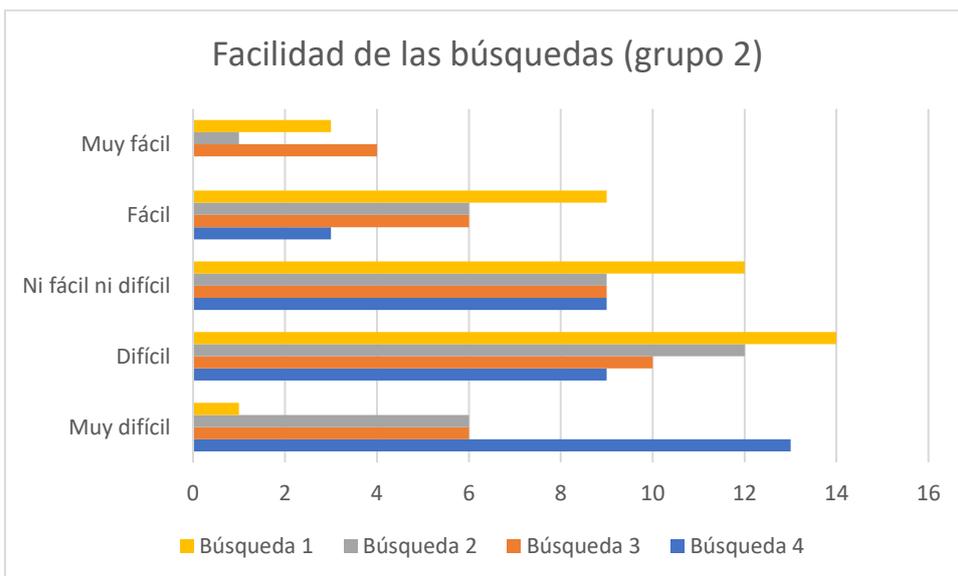


Gráfico 2: resultados de la encuesta sobre la facilidad de las búsquedas propuestas (grupo 2)

5.4.2.3 Otros resultados

Bajo estas líneas se representan gráficamente los resultados de las dos cuestiones añadidas al test de usabilidad en las que se pide al usuario expresar su grado de acuerdo con la afirmación

dada. De nuevo han sido medidos mediante una escala de Likert compuesta por cinco categorías: *totalmente de acuerdo*, *de acuerdo*, *ni de acuerdo ni en desacuerdo*, *en desacuerdo* y *totalmente en desacuerdo*. El eje vertical representa esta escala y el eje horizontal el número de usuarios que han manifestado su acuerdo o desacuerdo respecto de la afirmación.

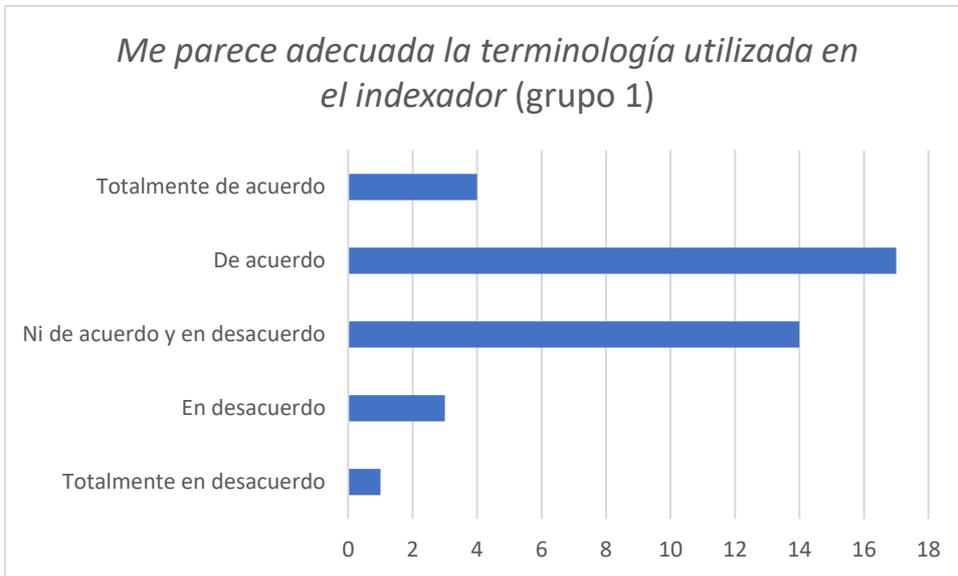


Gráfico 3: encuesta sobre la pertinencia de la terminología de ICAE (grupo 1)

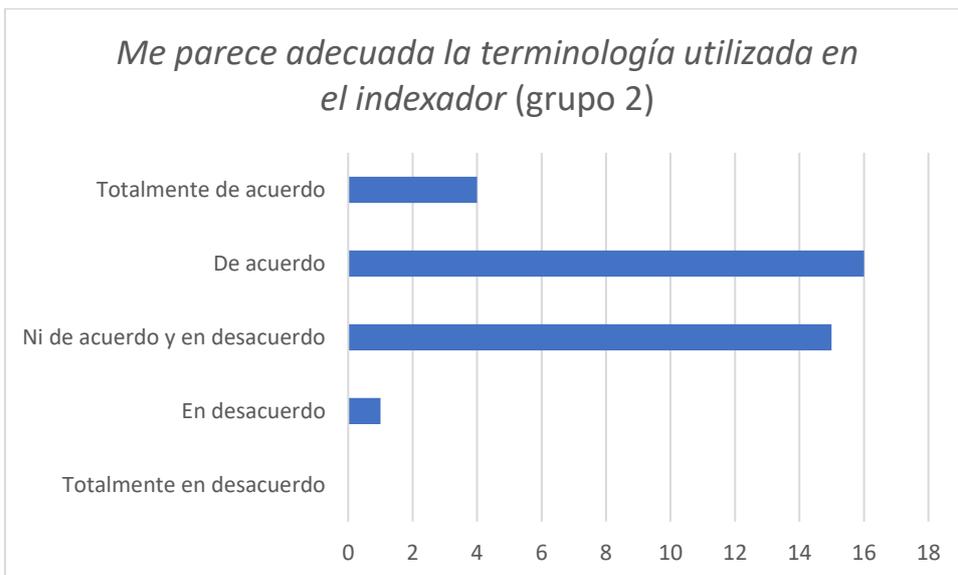


Gráfico 4: encuesta sobre la pertinencia de la terminología de ICAE (grupo 2)

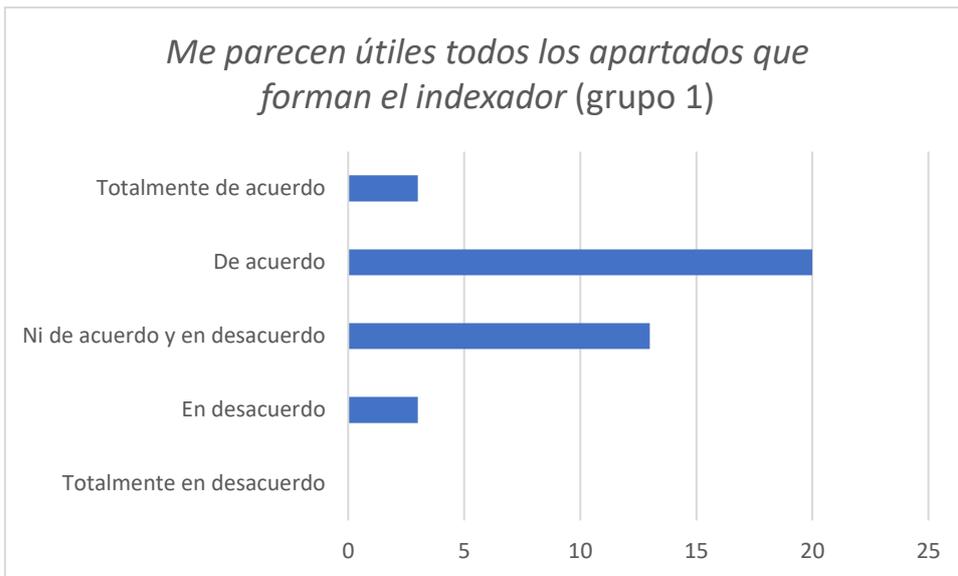


Gráfico 5: encuesta sobre la utilidad de los apartados de ICAE (grupo 1)

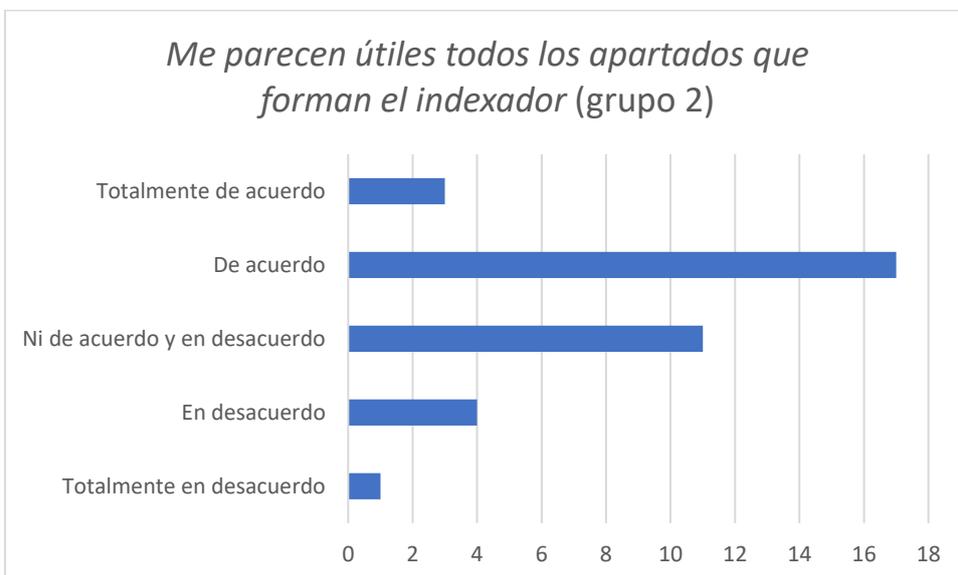


Gráfico 6: encuesta sobre la utilidad de los apartados de ICAE (grupo 2)

5.4.2.4 Resultados de la evaluación cualitativa

Los comentarios de los participantes en la evaluación se han clasificado mediante diez categorías distintas:

1. Formulario de búsqueda: abarca comentarios referentes a la confusión y dificultad que provoca un formulario de búsqueda con tantas categorías.
2. Navegación desapercibida: abarca comentarios que muestran que a los usuarios les ha pasado desapercibida la opción de la navegación como método de búsqueda más intuitivo.
3. Desconocimiento del sistema: abarca comentarios en los que los usuarios justifican la dificultad de las búsquedas por no estar familiarizados con el sistema ni tener tiempo suficiente para hacerlo durante la prueba.
4. Confusión de tipo de tarea y tipo de texto: abarca comentarios que expresan confusión entre *tipo de tarea* y *tipo de texto*.

5. Multivaluados: abarca comentarios relacionados con la dificultad de buscar valores de atributos multivaluados como *L1* o *Combinación de lenguas*.
6. Desplegables preferidos: abarca comentarios que expresan el deseo de que haya más metadatos de tipo controlado.
7. Recomendaciones de uso desapercibidas: abarca comentarios que muestran que a los usuarios les ha pasado desapercibida la sección de recomendaciones de uso.
8. Botón *intro* para la búsqueda: abarca comentarios que señalan como dificultad que las búsquedas no se lancen pulsando el botón *intro* del teclado y que sea necesario ir a pulsar *búsqueda* con el ratón.
9. Letra pequeña: abarca comentarios que expresan molestia ante un tamaño demasiado reducido en la tipografía del indexador.
10. Ventana taxonomía navegación: abarca comentarios en los que se señala como dificultad un tamaño demasiado pequeño de la taxonomía para la navegación que aparece a la izquierda.

Estas diez categorías se representan en el eje vertical de los gráficos siguientes. En el horizontal, se representan las veces que los usuarios han repetido comentarios respecto a dificultades, problemas o preferencias, de forma que se puede ver qué categorías se han repetido con más frecuencia en cada grupo.

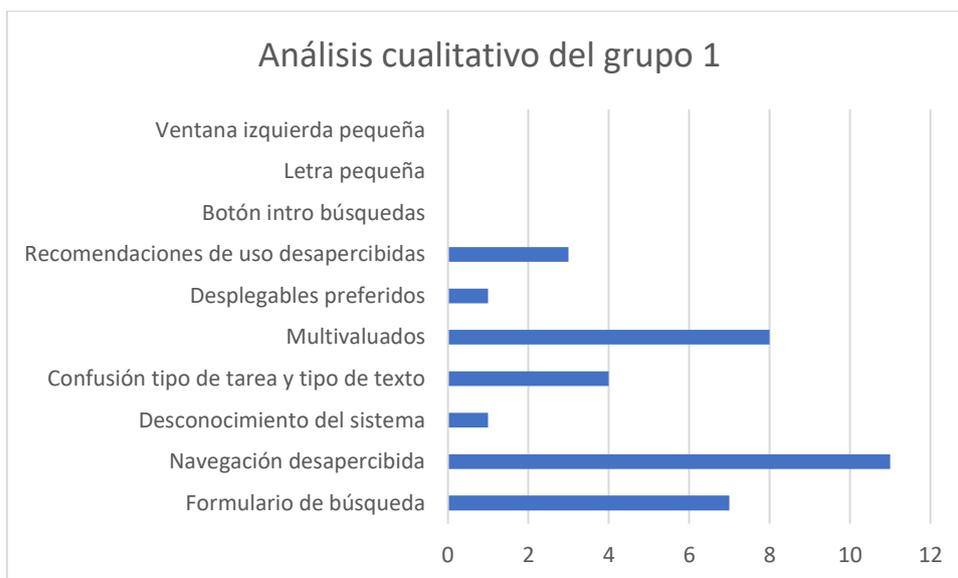


Gráfico 7: resultados del análisis cualitativo (grupo 1)

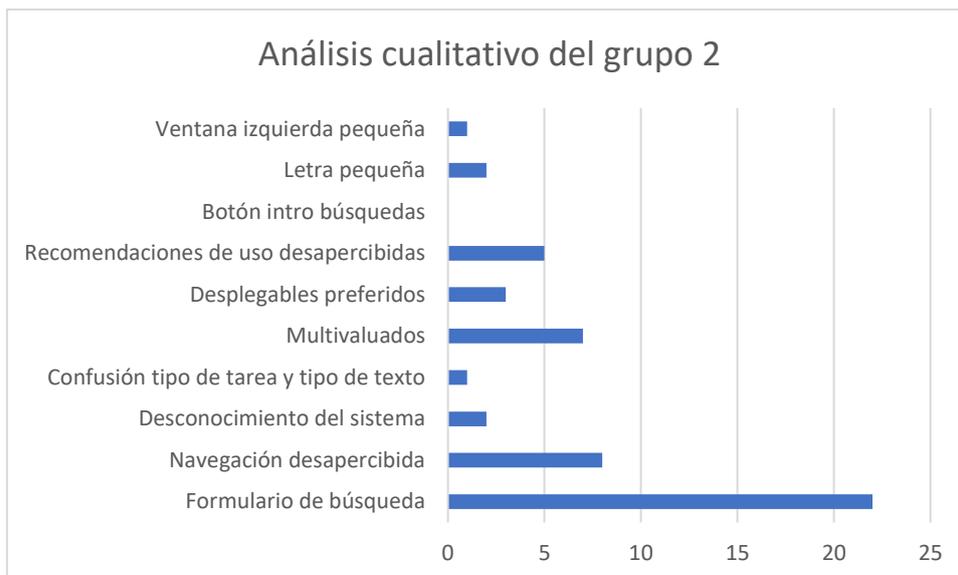


Gráfico 8: resultados del análisis cualitativo (grupo 2)

5.4.3 Discusión

El objetivo de evaluar el indexador es conocer sus puntos débiles y sus puntos fuertes y poder diseñar las mejoras necesarias. Esto se ha logrado mediante una evaluación deductiva de la calidad y una evaluación empírica de la usabilidad realizada por usuarios objetivo de ICAE.

Los resultados de la evaluación deductiva se consideran positivos teniendo en cuenta el modesto alcance del proyecto. Que se trate de un recurso de acceso abierto, con contenido académico de calidad y que haya sido y vaya a ser promocionado por la universidad son características que indican la calidad del recurso y que no son desdeñables sabiendo que ICAE es un indexador de tamaño reducido y desarrollado en el marco de una beca de colaboración.

Por exceder las posibilidades del trabajo, se han dejado por implementar la inclusión de un documento que exprese la política sobre el archivo en el repositorio y otras acciones de mercadotecnia.

Se han completado las tareas de evaluación con la evaluación empírica de la usabilidad. Los resultados presentados en los gráficos precedentes apuntan mayoritariamente a una baja usabilidad del indexador en el estado en el que se ha evaluado.

El grado F otorgado mediante el test de usabilidad tanto en el grupo 1 como en el grupo 2 es el primer síntoma de problemas en el sistema. No obstante, el Sistema de Escalas de Usabilidad no sirve para localizar los puntos de fricción (Sauro, 2016). Para ello se han propuesto al usuario una serie de situaciones que debía resolver y evaluar cualitativa y cuantitativamente mediante una pregunta abierta y una escala de Likert. Los resultados de estas evaluaciones explican los problemas principales del indexador.

En la representación del grado de dificultad que los usuarios han encontrado por búsqueda, se puede observar cómo han percibido la búsqueda 1 como menos dificultosa que la búsqueda 4. Esto es lógico, pues los ejercicios fueron planteados para que hubiese un incremento de la dificultad a medida que se iba avanzando. No obstante, es reseñable cómo las cuatro búsquedas han sido mayoritariamente calificadas por ambos grupos como *ni fácil ni difícil, difícil o muy difícil*.

Los resultados que se presentan en los gráficos 7 y 8 son los correspondientes a los datos cualitativos extraídos del cuestionario presentado a los usuarios, en el cual han dado cuenta de las dificultades que han percibido en el uso ICAE. Estos datos nos permiten apuntar con mayor precisión a los problemas que explican los resultados de la evaluación.

Como se ve en las representaciones gráficas, las categorías de dificultades más señaladas son *formulario de búsqueda*, *navegación desapercibida*, *recomendaciones de uso desapercibidas* y *multivaluados*. Es decir, los usuarios han criticado con firmeza la configuración del formulario de búsqueda, el problema de buscar varios valores para un mismo campo y la falta de una señalización más llamativa de secciones tan importantes para la recuperación de la información como son la navegación por categorías y las recomendaciones de uso.

Las críticas al formulario de búsqueda (ilustración 6) se deben a que la cantidad y variedad de campos de búsqueda que aparecen en él resulta excesiva y confusa para un usuario que entra por primera vez en el indexador y no está familiarizado con él. Este problema está relacionado con el hecho de que la navegación pasase generalmente desapercibida a los ojos de los usuarios, pues si el sistema de navegación hubiese sido más visible quizás no se habrían sentido tan frustrados ante una búsqueda compleja y desconocida. Así pues, es necesario aclarar todo lo posible al nuevo usuario que la búsqueda por formulario es una búsqueda avanzada y que no se debe acudir a ella en primera instancia sino a la navegación.

Los comentarios negativos sobre la dificultad de hacer una búsqueda con varios valores para un mismo campo (por ejemplo, la situación 4) están relacionados con el problema informático que ya se ha comentado sobre la imposibilidad de representar bien en OdA atributos multivaluados. Esta dificultad está relacionada con el cuarto problema que sacan a la luz los comentarios: la sección de recomendaciones también ha pasado desapercibida para el usuario. Esto ha provocado la desorientación de los usuarios especialmente al realizar la búsqueda de la situación 4. Aunque el problema informático no se puede resolver en esta etapa del proyecto, sí se puede resaltar lo máximo posible el apartado de recomendaciones de búsqueda, de forma que los posibles usuarios comprendan rápidamente de qué manera están representados estos campos cuando tienen varios valores y cómo se pueden recuperar objetos con estas características.

Los resultados positivos han sido aquellos relacionados con la afirmación sobre la terminología y la afirmación sobre la utilidad de los apartados. Como se puede ver en los gráficos 3, 4, 5 y 6, los usuarios han mostrado una mayor tendencia a estar de acuerdo con lo acertado de la terminología empleada en el indexador y con la utilidad de los apartados que lo forman. Lo primero es de gran relevancia viniendo de dos grupos de usuarios especializados en el estudio del español como lengua extranjera, pues indica que no es necesario hacer modificaciones en este sentido.

La evaluación empírica de la usabilidad que se ha realizado se considera adecuada dada la evidente utilidad que ha tenido para conocer los problemas del indexador. Además, la oportunidad de contar con las opiniones de usuarios que se encuentran dentro del objetivo del recurso como son estudiantes de enseñanza e investigación en ELE otorga un especial interés a los resultados de las encuestas.

Como punto débil de la propia evaluación hay que señalar que las circunstancias en que se recogieron los datos no fueron las idóneas, pues los usuarios se encontraban en su entorno universitario, había más de cuarenta personas en cada aula y, además, estaban cansados después de varias horas de clase. Ellos mismos explicaron que probablemente en un contexto de más calma y concentración habrían obtenido mejores resultados en las búsquedas y no

habrían tenido tanta sensación de dificultad. Sin embargo, esto no invalida unos resultados tan uniformes.

Tener las debilidades del indexador tan claramente señaladas permite trabajar en los cambios necesarios y explorar las posibilidades que hay para solventar las incidencias y mejorar la usabilidad de la aplicación. La implementación de estos cambios y mejoras se comentará en la siguiente sección.

5.5 Resultados de los cambios y mejoras

Los cambios que se han hecho en la implementación del indexador son aquellos necesarios para resolver en lo posible los problemas que se han podido deducir de las evaluaciones a las que se ha sometido al sistema y que han sido comentados en el apartado anterior. Estas modificaciones son las siguientes:

5.5.1 Modificaciones en los títulos de la barra de navegación

Como se puede apreciar en la ilustración 12, se ha sustituido *Navegación clasificada por EXPLORA, Buscador por Busca un objeto y Búsqueda por formulario por Refina tu búsqueda.*



Ilustración 12: nueva barra de navegación

5.5.2 Modificaciones en las recomendaciones de uso

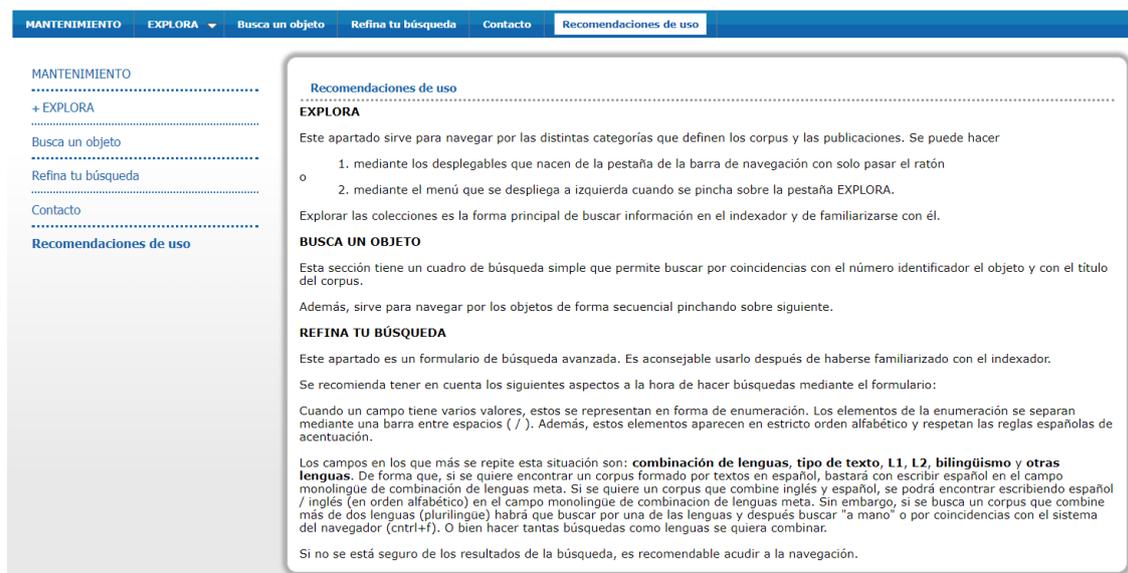


Ilustración 13: nueva sección de recomendaciones de uso

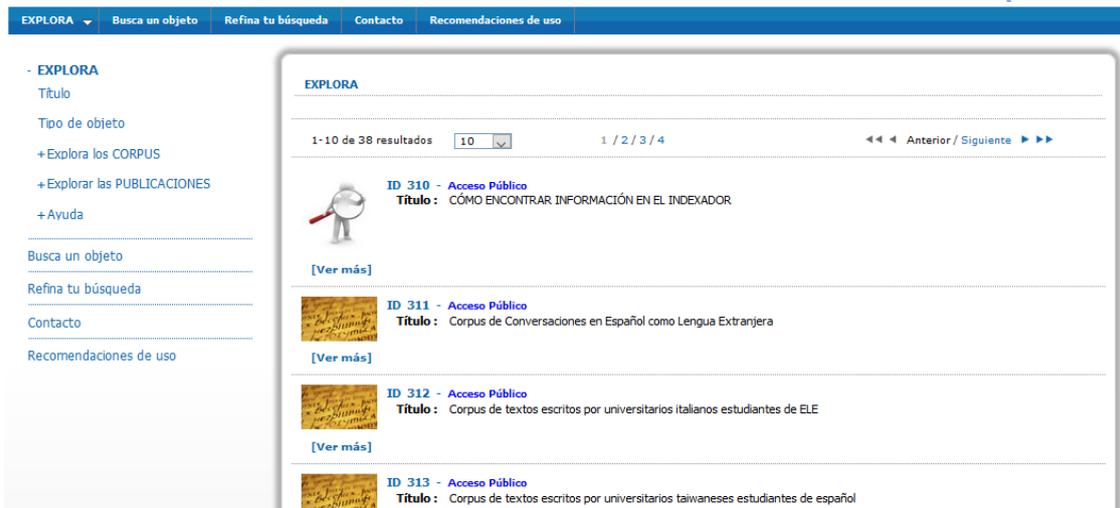


Ilustración 14: nuevo objeto digital de ayuda



Ilustración 15: nuevo objeto digital de ayuda

5.5.3 Discusión

Los resultados de las evaluaciones han hecho evidente la necesidad de llevar a cabo ciertos cambios en el indexador. Estos se han realizado con el objeto de mejorar el sistema y así acercarse más al objetivo general de terminar con la diseminación de corpus de aprendices de español por internet.

Para mejorar la experiencia de usuario, se han hecho cambios en la nomenclatura de las categorías que forman la barra de navegación. Se ha sustituido *Navegación clasificada* por *EXPLORA*, *Buscador* por *Busca un objeto* y *Búsqueda por formulario* por *Refina tu búsqueda*. Esta medida se ha tomado con el fin de evitar que la navegación pasase desapercibida y que los

usuarios acudieran en primer lugar a la pestaña antes llamada *Búsqueda por formulario*. Así, se ha resaltado la navegación y se ha indicado que la búsqueda por formulario es un tipo de búsqueda avanzada y que no se debe acudir a ella si no se está familiarizado con la aplicación y la representación que en ella se hace de la información.

La segunda medida tomada ha sido la de resaltar y ampliar la sección de *Recomendaciones de uso* y ayudar así a saber cómo se representan los atributos multivaluados. Para ello, se ha convertido el primer objeto digital del conjunto en un objeto de ayuda que contiene el mismo texto que la sección *Recomendaciones de uso*. El texto con las recomendaciones, además, se ha enriquecido con nuevas explicaciones. Esta se considera una solución acertada porque dicho objeto es el primero que aparece tanto cuando se accede a la sección *Explora* como cuando se entra en *Busca un objeto*. De esta manera hay mucha más probabilidad de que el usuario se encuentre con estas indicaciones que le orientarán en el uso de ICAE.

Sin embargo, se han dejado por atender algunos problemas por exceder las posibilidades de esta etapa del proyecto. Aun así, se quieren explicar aquí algunas soluciones posibles para el futuro. Por ejemplo, la confusión entre *tipo de texto* y *tipo de tarea* se podría evitar convirtiendo el tipo de dato de *tipo de texto* en *clasificado* en lugar de *texto*.

Por otro lado, se podría haber profundizado más en la mejora del sistema de búsquedas cambiando el campo *Título* por *Descripción*, de forma que se pudieran recuperar los objetos por palabras clave incluidas en esta descripción, pues este es el campo fijo que sirve para hacer las búsquedas por coincidencia en la sección *Busca un objeto*. Este cambio conllevaría una notable mejora en el sistema, pero la estimación de tiempo que requiere establecer una serie de palabras clave descriptivas, elaborar una descripción para cada corpus o publicación, modificar la configuración del esquema de datos, cambiar cada objeto y explicar el cambio en las recomendaciones de uso supera las posibilidades del presente trabajo.

Otras quejas y peticiones de los evaluadores como el tamaño de la tipografía y del menú de la izquierda, la imposibilidad de buscar a la vez en el formulario varios valores de un mismo campo o el funcionamiento del botón *intro* del teclado en OdA son problemas informáticos. Es decir, para resolverlos sería necesaria la participación de un técnico, recurso del que carece este proyecto hasta el momento.

A pesar de no haber implementado todos los cambios posibles, se considera que las modificaciones llevadas a cabo cumplen con el objetivo de mejorar el sistema basándose en los comentarios de los usuarios, pues como se ha dicho, resaltan la navegación y las recomendaciones de uso. No obstante, para saberlo con seguridad habría que realizar de nuevo una prueba empírica a esta última versión del sistema. Volver a evaluar la usabilidad mediante la Escala de Sistemas de Usabilidad permitiría saber si las mejoras se traducen en un mayor éxito en comparación con la versión anterior, pues es un test especialmente concebido para dar un resultado global que pueda ser comparado con otras soluciones y elegir así la de mayor nivel de usabilidad. Esto solo se puede dejar aquí planteado como siguiente paso, pues escapa al alcance de esta etapa del proyecto.

6 Conclusiones y trabajo futuro

Con el objetivo de dar visibilidad y facilitar la recuperación de los corpus de aprendices de español existentes, se ha diseñado y desarrollado un prototipo de indexador para comprobar la hipótesis de que un sistema así es especialmente adecuado para representar información sobre

corpus. Los resultados de esta tarea han sido posteriormente evaluados y, como último paso, se han incluido una serie de mejoras basadas en las conclusiones de la evaluación.

El desarrollo del Indexador de Corpus de Aprendices de Español ha permitido la centralización y organización del acceso y la información de 46 corpus de aprendices de español como lengua meta y de una publicación basada en un corpus. Se ha creado un sistema que recoge un buen número de datos organizados, que es capaz de enlazar a recursos externos y también alojarlos en su propio servidor y que puede ser modificado y ampliado fácilmente. Además, facilita la recuperación de la información y los recursos mediante un sistema de navegación a través de un esquema de facetas y subfacetas formada con los metadatos de los objetos y mediante dos sistemas de búsqueda diferentes.

Constituye, por lo tanto, un doble beneficio para la comunidad académica: por un lado, los corpus son más fácilmente localizables y recuperables; por otro, los investigadores que tengan un corpus, tanto si está ya publicado como si no, tienen un sistema mediante el cual compartirlo con la comunidad. Se puede afirmar, por lo tanto, que se ha cumplido con el objetivo general perseguido por este proyecto: resolver el problema generado por la dispersión de materiales e información de los corpus mencionados.

No obstante, el indexador en su estado actual tiene varios puntos débiles. Como problema informático, resalta especialmente la dificultad de representar campos que pueden tener varios valores, como es el caso de la lengua materna de un grupo de informantes. Además, hay otros aspectos que dificultan la usabilidad señalados por los evaluadores, como el tamaño demasiado reducido de la tipografía y del componente que contiene la taxonomía de navegación. En cuanto a problemas de diseño, hay que señalar el desaprovechamiento actual del sistema de búsqueda simple por un error de concepción del sistema, que sería mucho más útil si buscarse entre las descripciones de cada objeto y no en el título.

Estas debilidades que tiene el sistema se pueden solucionar. Los primeros, con la colaboración de un profesional de la informática; y los segundos, incluyendo las mejoras que se sugieren en este documento. Los problemas informáticos se derivan de la elección de OdA para implementar el sistema y los de diseño, de errores iniciales de concepción. No se considera, por tanto, que supongan una contradicción con la afirmación de que un repositorio o indexador es la mejor manera de las que se han encontrado para recoger la información relacionada con corpus de aprendices de español.

Se considera que los puntos fuertes del indexador aportan más de lo que restan los débiles. Especialmente porque los problemas que se han encontrado hasta el momento, como se ha explicado, se pueden resolver.

En comparación con los sistemas revisados que no son repositorios o indexadores y que se encargan de problemas similares, ICAE contiene más información y, sobre todo, alberga datos sobre una cantidad de corpus de español que el resto de los recursos no incluye. También ofrece unos métodos de navegación y búsqueda, en general, más completos que los correspondientes a estos otros sistemas.

Por lo expuesto en los párrafos anteriores, se considera que se valida la hipótesis defendida durante el proyecto: de todas las formas en que se podría recoger y ordenar la información relativa a corpus de aprendientes de español, la más adecuada es un repositorio o indexador. La principal conclusión es, por tanto, que el camino iniciado con el desarrollo del prototipo de indexador es una vía que merece la pena continuar.

La evaluación de la nueva versión del indexador, así como las modificaciones explicadas que no ha sido posible implementar quedan como parte del futuro trabajo que se puede hacer con este sistema. Al tratarse de un proyecto evolutivo, no tiene por qué terminar aquí. Además de los cambios, será necesario continuar añadiendo nuevos corpus que hayan pasado desapercibidos o que no existan aún.

Otra tarea futura es la de la implementación y mejora de la colección de publicaciones relacionadas con los corpus, de la cual solo existe el diseño y un prototipo en la actualidad.

Por último, el Indexador de Corpus de Aprendices de Español podrá ser enriquecido con nuevas facetas que no han sido diseñadas en esta fase del trabajo. Sería muy interesante, por ejemplo, describir los corpus mediante facetas relacionadas con el análisis de los errores que contiene cada corpus o con los recorridos didácticos que se pueden idear a partir de ellos.

7 Bibliografía

AA. VV. (2008). «Adquisición de segundas lenguas». *Diccionario de términos clave de ELE*.

Recuperado de:

https://cvc.cervantes.es/ensenanza/biblioteca_ele/diccio_ele/diccionario/adquisicion.htm

AA. VV. (2008). «Lengua meta». *Diccionario de términos clave de ELE*. Recuperado de: https://cvc.cervantes.es/ensenanza/biblioteca_ele/diccio_ele/diccionario/lenguameta.htm

Calvo-Fernández Rodríguez, A., Ortega Santamaría, S., Valls Sáez, A. (Desconocido) *Métodos de evaluación con usuarios*. Consultado el 5-10-2018. Recuperado de

http://openaccess.uoc.edu/webapps/o2/bitstream/10609/9861/4/PID_00176614.pdf

Centre for English Corpus Linguistics. Learner Corpora around the World. Louvain-la-Neuve: Université catholique de Louvain. Consultada el 02-10-2018 en <https://uclouvain.be/en/research-institutes/ilc/cecl/learner-corpora-around-the-world.html>

Chandler, C., Hearst, M. A. y Smalley, P. (2006). *Faceted Metadata for Information Architecture and Search*. En CHI 2006 Course, Montreal, Canadá. Recuperado de http://bailando.sims.berkeley.edu/talks/chi_course06.pdf

CHILDES Corpora. Child Language Data Exchange System. <https://childes.talkbank.org/access/>

Colecciones de objetos digitales: visión general. Material de formación aportado por Ana Fernández-Pampillón, marzo de 2018

Corpus de aprendices de español (CAES). Consultado el 5-12-2017 en http://www.cervantes.es/lengua_y_ensenanza/tecnologia_espanol/caes.htm

Corpus Resource Database (CoRD). VARIENG. <http://www.helsinki.fi/varieng/CoRD/index.html>

De Giusti, M. R. (2017, octubre). *Indicadores de calidad en repositorios de acceso abierto*. Ponencia presentada en la Semana de Acceso Abierto, La Plata, Argentina. Recuperado de http://sedici.unlp.edu.ar/bitstream/handle/10915/63176/Presentaci%C3%B3n_diapositivas_.pdf-PDFA.pdf?sequence=3

Devin, F. (2017, 25 de febrero). *Sistemas de Escalas de Usabilidad: ¿qué es y para qué sirve? UXpañol*. Recuperado de <http://uxpanol.com/teoria/sistema-de-escalas-de-usabilidad-que-es-y-para-que-sirve/>

Evaluación de bibliotecas. Material de formación aportado por Almudena Caballos, noviembre de 2018

Faya Ornia, G. (2015) Propuesta de clasificación de corpus textuales En Álvarez Álvarez, S. (coord.), Arnáiz Uzquiza, V. (coord.), Fernández Muñiz, R. (coord.), Ortego Antón, M. T., Sánchez Nieto, M. T. (coord.) y Santamaría Ciordia, L. (coord.), *Metodologías y aplicaciones en la investigación en traducción e interpretación con corpus* (pp. 339-356). España: Universidad de Valladolid. Recuperado de <http://uvadoc.uva.es/handle/10324/16449>

Fernández-Pampillón Cesteros, A., Fernández-Valmayor Crespo, A. y Varadero Software Factory, VSF (2013). *Guía de Gestión del repositorio de Objetos Digitales OdA*. [] (No publicado). Recuperado de <https://eprints.ucm.es/20263/>

Fernández-Pampillón Cesteros, A. (2017). The Role of E-Vocabularies in the Description and Retrieval of Digital Educational Resources. *Education Sciences*, 7(1), 33. doi: 10.3390/educsci7010033. Recuperado de <http://www.mdpi.com/2227-7102/7/1/33/htm>

Fernández-Pampillón, A., Gayoso, J., Sarasa, A. y Sierra, J. L. (s.f.). *De OdA a Clavy, dos aplicaciones orientadas a la creación de repositorios de objetos digitales académicos*. Monográfico de la I Jornadas UCM. La investigación en Ciencias Sociales y Humanidades. Desafíos y perspectivas de entornos digitales (en prensa)

Fevin, F. (2017). *Sistema de Escalas de Usabilidad. ¿Qué es y para qué sirve?* En UXpañol, www.uxpanol.com. Recuperado de <http://uxpanol.com/teoria/sistema-de-escalas-de-usabilidad-que-es-y-para-que-sirve/>

García, N. E. y Jaroszczuk, S. E. (2009). Objetos digitales: una experiencia de representación con metadatos Dublin Core. En Biblioteca Nacional de Buenos Aires (Ed.), *Experiencias en la organización y tratamiento de la información en las bibliotecas argentinas* (pp. 193-206). Buenos Aires: Biblioteca Nacional. Recuperado de https://www.bn.gov.ar/resources/conferences/pdfs/ponencia_garcia_jaroszczuk.pdf

Granger, S. (2008). Learner corpora. En Lüdeling, A. y Kytö, M. (eds.), *Corpus Linguistics. An International 5 Handbook. Volume 1* (pp. 259-275). Berlin & New York: Walter de Gruyter.

Granger, S. y Paquot, M. (2017, diciembre). *Towards standardization of metadata for L2 corpora*. Ponencia presentada en el CLARIN workshop on Interoperability of Second Language Resources and Tools, Gotemburgo, Suecia. Recuperado de https://sweclarin.se/sites/sweclarin.se/files/event_atachements/Granger_Paquot_Metadata_G%C3%B6teborg_final.pdf

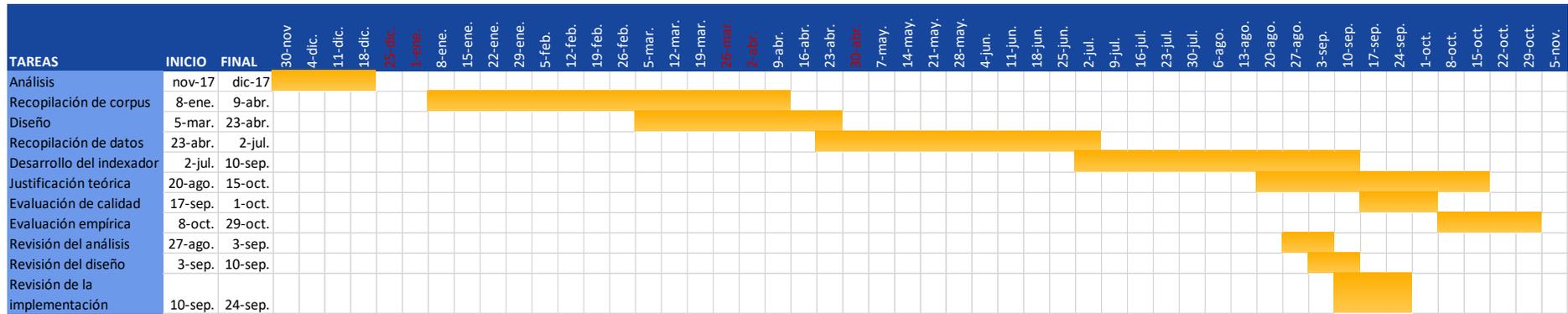
Hearst, M. A. (2006). Design recommendations for hierarchical faceted search interfaces. *Proc. SIGIR 2006, Workshop on Faceted Search* (pp. 26-30). Recuperado de <http://flamenco.berkeley.edu/papers/faceted-workshop06.pdf>

Hearst, M. A., (2006). Clustering versus faceted categories for information exploration. *Communications of the ACM*, 49(4), 59-61. doi: 10.1145/1121949.1121983. Recuperado de <https://dl.acm.org/citation.cfm?doid=1121949.1121983>

Hearst, M. A., Li, K., Swearingen, K. y Yee, K. P. Faceted Metadata for Image Search and Browsing (2003). En *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 401-408). doi: 10.1145/642611.642681. Recuperado de <http://flamenco.berkeley.edu/papers/flamenco-chi03.pdf>

- HZSK Repository. (2018). <https://corpora.uni-hamburg.de/hzsk/en/repository-search>
- IEEE Std. 830-1998 830-1998 - *IEEE Recommended Practice for Software Requirements Specifications*, 1998.
- Krug, B. (2017). An Overview of the System Usability Scale in Library Website and System Usability Testing. En *Weabe. Journal OF Library User Experience. Volumen 1. Número 6*. Recuperado de <https://quod.lib.umich.edu/w/weave/12535642.0001.602?view=text;rgn=main>
- La catalogación de objetos digitales mediante metadatos*. Material de formación aportado por Ana Fernández-Pampillón, marzo de 2018
- Martinez-Arellano, F. y Palma, J. (2017, septiembre). *Metadatos para la organización de publicaciones académicas*. Ponencia presentada en el 9º Encuentro de Catalogación y Metadatos, México. doi: 10.13140/RG.2.2.16768.53767. Recuperado de https://www.researchgate.net/publication/323737520_Metadatos_para_la_organizacion_de_publicaciones_academicas
- Online/Downloadable Corpora* (s.f.). Recuperado de https://www.leuphana.de/fileadmin/user_upload/PERSONALPAGES/_abcd/barron_anne/files/Hyperlink_-_Freely_available_corpora-updated.pdf
- Pressman, R. S. (2010). *Ingeniería del software. Un enfoque Práctico* (7ª ed.). México: McGraw-Hill.
- Sauro, J. (2016, 2 de mayo). Measuring Usability With The System Usability Scale (SUS). *User Focus*. Recuperado de <https://www.userfocus.co.uk/articles/measuring-usability-with-the-SUS.html>
- Sauro, J., Lewis, J. (2016). *Quantifying the user experience: Practical statistics for user research*. Amsterdam; Waltham, MA: Elsevier/Morgan Kaufmann.
- Tunkelang, D. (2009). *Faceted search*. Estados Unidos: Morgan & Claypool. Recuperado de <http://www.iro.umontreal.ca/~nie/IFT6255/Books/FacetedSearch.pdf>
- Weisser, M. (2016). *Corpus-based Linguistics Links*. http://martinweisser.org/corpora_site/CBLLinks.html

Anexo I: Diagrama de Gantt



Anexo II: Especificación de requisitos

Introducción

El presente documento pretende constituir la especificación de los requisitos de un indexador de corpus de aprendices de español. Es decir, compuestos por textos (en un sentido amplio) generados por estudiantes de español como lengua extranjera.

Propósito

La especificación de requisitos tiene por objetivo definir las especificaciones funcionales, no funcionales y del sistema para la implementación de un sistema de información.

Ámbito del sistema

El Indexador de Corpus de Aprendices de Español contendrá información relativa a los diferentes corpus lingüísticos existentes de muestras de lengua oral y/o escrita producidas por hablantes no nativos de español, es decir, personas que lo han aprendido o lo estudian como segunda lengua (L2) o lengua extranjera (LE). También podrá albergar datos sobre corpus que aún están siendo elaborados.

El indexador será implementado basándose en el modelo jerárquico facetado que utiliza la herramienta OdA. El sistema permitirá el acceso a estos datos mediante un método de búsqueda avanzada que se sustentará en el esquema de datos y metadatos y la navegación por los campos que se definan. Se podrán hacer consultas cruzando criterios.

El sistema debe contener información sobre los siguientes aspectos:

- De los corpus:
 - título,
 - dirección web,
 - fuente de información (especialmente cuando el corpus no está en línea),
 - tipo de corpus,
 - formato,
 - tamaño,
 - objetivos,
 - forma de acceso,
 - combinación de lenguas,
 - tipo de tarea
 - tipo de texto,
 - fuente,
 - fecha de obtención,
 - fecha de publicación,
 - otros rasgos (anotación, por ejemplo);
 - número de informantes,
 - publicaciones relacionadas;
- del creador o creadores:
 - autoría y créditos,
 - editor,
 - contacto,
 - rol del investigador;

- de los informantes:
 - L1,
 - bilingüismo,
 - L2,
 - conocimiento de otras lenguas,
 - edad,
 - nivel de español,
 - nivel de estudios,
 - género,
 - profesión,
 - nivel sociocultural.

Es necesario, además, que el indexador se pueda ampliar con la introducción tanto de nuevos corpus y datos que surjan en el futuro como de nuevos campos o facetas que supongan una mejora del sistema.

Algunos ejemplos de los tipos de consultas que se han de poder hacer son:

- Búsqueda por L1
- Corpus bilingües/corpus multilingües (y dentro de estos por combinaciones de lenguas)
- Búsqueda por lengua oral/lengua escrita. ¿Qué bases de datos son de lengua oral? ¿Cuáles son de lengua escrita? ¿Cuáles combinan ambas (si las hay)?
- Búsqueda por géneros textuales/discursivos (relatos, descripciones, argumentaciones, exposiciones, conversaciones...).
- En el caso de la lengua oral, búsqueda por tipo de datos (grabaciones en audio, grabaciones en vídeo, transcripciones escritas...).
- Búsqueda por edades (niños/adultos).
- Búsqueda por sexo.

Además, es necesario que en las consultas se puedan cruzar estos criterios.

Se prevé que los usuarios de ICAE sean estudiantes de máster y doctorado que investiguen sobre cuestiones de español como segunda lengua, así como profesores e investigadores. Por el momento, habrá dos tipos de acceso, el de superadministrador (con competencias como modificar el esquema y la información que recoge el indexador) y el de visitante ocasional (que solo podrá consultarla).

El objetivo de este proyecto es recopilar información sobre los mencionados corpus. El hecho de tener acceso a esta información ordenada y centralizada en un mismo lugar supondrá un beneficio para los investigadores en adquisición de L2 y, en general, para la comunidad académica y docente interesada en el español como lengua extranjera.

No existen documentos de nivel superior que referenciar para este sistema.

Definiciones, acrónimos y abreviaturas

ELE	Español como lengua extranjera
L1	Lengua materna
L2	Segunda lengua
LE	Lengua extranjera

Corpus	En lingüística, un conjunto, más o menos amplio, de textos en formato electrónico que han sido reunidos en una aplicación informática, según un determinado diseño, para facilitar el estudio de la lengua o variedad lingüística de la que esos textos han sido extraídos.
Corpus de aprendices	Corpus que contienen textos producidos por personas que están aprendiendo una determinada lengua con diferentes lenguas iniciales o familiares [...] y distintos grados de conocimiento de la lengua objeto. ¹⁷
Indexador	Recurso que centraliza información dispersa.
OdA	«Objetos digitales de Aprendizaje: Aplicación web diseñada para crear y mantener colecciones de Objetos Digitales, inicialmente de Aprendizaje, pero actualmente de cualquier tipo». ¹⁸

Referencias

Standard IEEE 830 - 1998	IEEE
CAES (Corpus de aprendices de español)	http://www.cervantes.es/lengua_y_ensenanza/tecnologia_espanol/caes.htm
Guía de gestión OdA 2.5	https://eprints.ucm.es/20263/12/GUIA_ODA_v2_5%20%282%29.pdf

Visión general del documento

Este documento está compuesto por una introducción al mismo, una descripción donde se da cuenta de los factores generales que afectan al producto y sus requisitos y una descripción detallada de los requisitos específicos del sistema.

¹⁷ Definiciones extraídas de

http://www.cervantes.es/lengua_y_ensenanza/tecnologia_espanol/caes.htm .

¹⁸ Guía de Gestión de OdA 2.5, https://eprints.ucm.es/20263/12/GUIA_ODA_v2_5%20%282%29.pdf .

Descripción general

Perspectiva del producto

El producto objeto de esta especificación es dependiente de la aplicación *software* OdA, pues está creado a partir de las posibilidades de esta herramienta.

Funciones del producto

- Realización de búsquedas.
- Navegación por los distintos campos.
- Administración de usuarios.
- Administración de roles.
- Adición de datos.
- Adición de campos.
- Adición de objetos digitales (corpus).
- Modificación del esquema de datos.
- Modificación o borrado de datos y de objetos digitales.

Características de los usuarios

Los usuarios de este sistema serán mayoritariamente estudiantes de máster y doctorado que investiguen cuestiones relacionadas con el español como lengua meta y profesores e investigadores. También podría ser útil a otros profesionales de ELE así como a estudiantes de Filología Hispánica o carreras similares.

Restricciones

Las restricciones respecto al desarrollo y diseño del sistema vienen impuestas por la herramienta con que se va a elaborar: OdA. La documentación de esta aplicación se puede consultar en http://oda-fec.org/cocle/bo/download/475/GUIA_ODA_2013-v7.pdf.

Suposiciones y dependencias

La implementación del producto como conjunto de colecciones en el repositorio OdA puede verse afectada por los cambios a los que esté sujeto dicho repositorio. El recurso objeto de este proyecto será totalmente dependiente de la aplicación donde va a estar alojado, de forma que está limitado a sus posibilidades y sujeto a sus cambios.

Requisitos futuros

En el futuro se podrían incluir varias mejoras en el sistema, a saber:

- Añadir facetas al indexador (por ejemplo, recorridos didácticos o análisis de errores).
- Añadir, modificar y borrar datos o metadatos.
- Añadir, modificar y borrar campos de los esquemas de datos o metadatos.
- Hacer uso de otros roles de los otros roles de usuario que facilita OdA (administrador y usuario).

Requisitos específicos

Interfaces externas

El sistema ICAE estará almacenado en OdA y, por tanto, utilizará la interfaz de este *software*.

Funciones

Requisito funcional 1

Código	Nombre	Fecha	Grado de necesidad	
RF1	Realización de búsquedas	05/07/2018	Esencial	
Descripción	Los datos del ICAE deben poderse consultar a través de búsquedas tanto simples como complejas.			
Entradas	Fuente	Salida	Destino	Restricciones
Seleccionar opciones en los campos del formulario de búsqueda	Formulario de búsqueda generado en OdA a partir del esquema de datos.	Mostrar en pantalla los datos resultantes de la consulta.	Usuario	
Proceso	No aplica			
Efecto colateral	No aplica			

Requisito funcional 2

Código	Nombre	Fecha	Grado de necesidad	
RF2	Navegación por el esquema de datos.	05/07/2018	Esencial	
Descripción	El sistema debe permitir al usuario navegar por los distintos campos que forman los esquemas de datos, metadatos y datos de los recursos.			
Entradas	Fuente	Salida	Destino	Restricciones
Botón de «navegación clasificada».	Datos seleccionados para la navegación de los esquemas de datos, metadatos y datos de los recursos.	Despliegue del esquema de navegación.	Usuario	
Proceso	No aplica			
Efecto colateral	No aplica			

Requisito funcional 3

Código	Nombre	Fecha	Grado de necesidad	
RF3	Creación de usuarios	05/07/2018	Esencial	
Descripción	El sistema debe permitir la creación de diferentes usuarios para asignarles permisos de acuerdo con sus tareas.			
Entradas	Fuente	Salida	Destino	Restricciones
Datos del usuario: nombre, apellidos, identificador y clave de acceso, correo y rol.	Formulario de ingreso de datos.	Usuario con acceso al sistema.	Base de datos.	Para gestionar usuarios hay que estar conectado como superadministrador.
Proceso	En el menú <i>Mantenimiento > Usuario</i> , pulsando el botón «nuevo» aparece un formulario en el que rellenar los campos citados.			
Efecto colateral	No aplica			

Requisito funcional 4

Código	Nombre	Fecha	Grado de necesidad	
RF4	Modificación de usuarios	05/07/2018	Esencial	
Descripción	El sistema debe permitir los datos de los usuarios.			
Entradas	Fuente	Salida	Destino	Restricciones
Datos del usuario: nombre, apellidos, identificador y clave de acceso, correo y rol.	Formulario de ingreso de datos	Datos de usuario modificados	Base de datos	Hay que estar conectado como superadministrador
Proceso	El superadministrador del sistema podrá editar los datos pulsando sobre el icono de editar asignado al nombre del usuario concreto en el menú <i>Mantenimiento > Usuario</i> .			
Efecto colateral	No aplica.			

Requisito funcional 5

Código	Nombre	Fecha	Grado de necesidad	
RF5	Eliminación de usuarios	05/07/2018	Esencial	
Descripción	El sistema debe permitir eliminar usuarios.			
Entradas	Fuente	Salida	Destino	Restricciones
Identificación de usuario	Formulario de ingreso de datos	Confirmación por pantalla	Propiedad del usuario modificada	Hay que estar identificado como superadministrador para realizar esta acción.
Proceso	El administrador del sistema tendrá una opción en el menú <i>Gestión de usuarios</i> que le permitirá eliminar a usuarios concreto.			
Efecto colateral	Los usuarios eliminados pasarán a ser visitantes ocasionales.			

Requisito funcional 6

Código	Nombre	Fecha	Grado de necesidad	
RF6	Modificar un objeto digital	05/07/2018	Esencial	
Descripción	El sistema debe permitir la modificación, adición y eliminación de datos, metadatos y recursos a los objetos digitales.			
Entradas	Fuente	Salida	Destino	Restricciones
Formulario editable de objeto digital.	Formulario.	Ficha editada del objeto digital.	Base de datos.	Estar identificado como superadministrador
Proceso	El superadministrador del sistema tendrá la opción de acceder al formulario para editar los objetos digitales pulsando sobre el icono de editar perteneciente a cada objeto o sobre el ID del mismo.			
Efecto colateral	No aplica.			

Requisito funcional 7

Código	Nombre	Fecha	Grado de necesidad	
RF7	Añadir campos al modelo de datos.	05/07/2018	Esencial	

Descripción	El sistema debe permitir incluir nuevos campos en el modelo de datos.			
Entradas	Fuente	Salida	Destino	Restricciones
Rellenar campos del formulario.	Formulario para añadir nuevo campo al modelo.	Modelo de datos con un nuevo campo.	Base de datos.	Estar conectado como superadministrador
Proceso	El superadministrador podrá acceder al formulario para añadir un nuevo campo al modelo de datos pulsando sobre el botón «nuevo» en la sección <i>objetos digitales>modelo de datos</i> .			
Efecto colateral				

Requisito funcional 8

Código	Nombre	Fecha	Grado de necesidad	
RF8	Modificar campos del modelo de datos.	05/07/2018	Esencial	
Descripción	El sistema debe permitir modificar las características de los campos que conforman el modelo de datos.			
Entradas	Fuente	Salida	Destino	Restricciones
Formulario editable del campo.	Formulario de campo del modelo de datos.	Campo modificado.	Base de datos.	Estar identificado como superadministrador. La aplicación OdA no permite modificar el tipo de dato correspondiente a un campo.
Proceso	El superadministrador podrá acceder al formulario para editar las características de los campos ya creados pulsando sobre el nombre del campo o el icono de editar correspondiente.			
Efecto colateral	No aplica.			

Requisito funcional 9

Código	Nombre	Fecha	Grado de necesidad	
RF9	Borrar campos del modelo de datos	05/07/2018	Esencial	
Descripción	El sistema debe permitir la eliminación de campos del modelo de datos.			
Entradas	Fuente	Salida	Destino	Restricciones
Botón de eliminar.	Formulario de campo del modelo de datos.	Campo eliminado del modelo.	Base de datos.	Estar identificado como superadministrador
Proceso	El superadministrador podrá eliminar campos que forman el modelo de datos pulsando el botón de eliminar dentro del formulario editable correspondiente al campo.			
Efecto colateral	Borrar los valores atribuidos al campo o atributo que haya sido borrado en todos los objetos digitales de la colección.			

Requisito funcional 10

Código	Nombre	Fecha	Grado de necesidad	
RF10	Añadir objetos digitales	05/07/2018	Esencial	
Descripción	El sistema debe permitir la adición de objetos digitales a la colección.			
Entradas	Fuente	Salida	Destino	Restricciones
Botón «nuevo» en la sección «objetos digitales».	Formulario	Nuevo objeto digital	Colección de objetos digitales del indexador.	Estar identificado como administrador o superadministrador
Proceso	Administrador y superadministrador podrán añadir nuevos objetos pulsando sobre el botón «nuevo» de la sección <i>objetos digitales>objetos</i> .			
Efecto colateral	No aplica.			

Requisitos de rendimiento

No se especifican requisitos de rendimiento.

Requisitos de datos

No se especifican requisitos de datos.

Restricciones de diseño

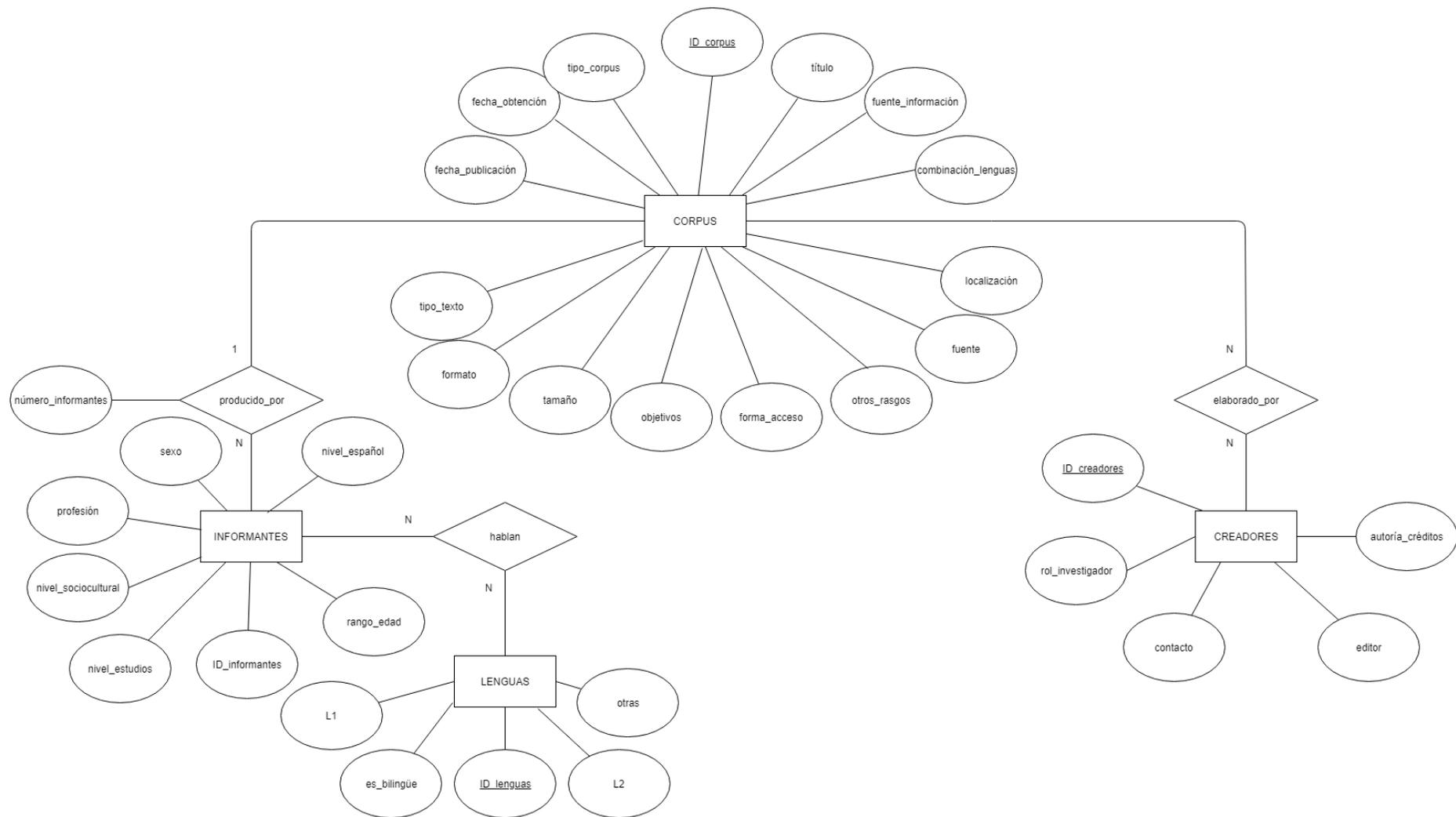
Las restricciones en el diseño del indexador serán las impuestas por la herramienta que usará para implementar el sistema: OdA.

Atributos del sistema

Seguridad

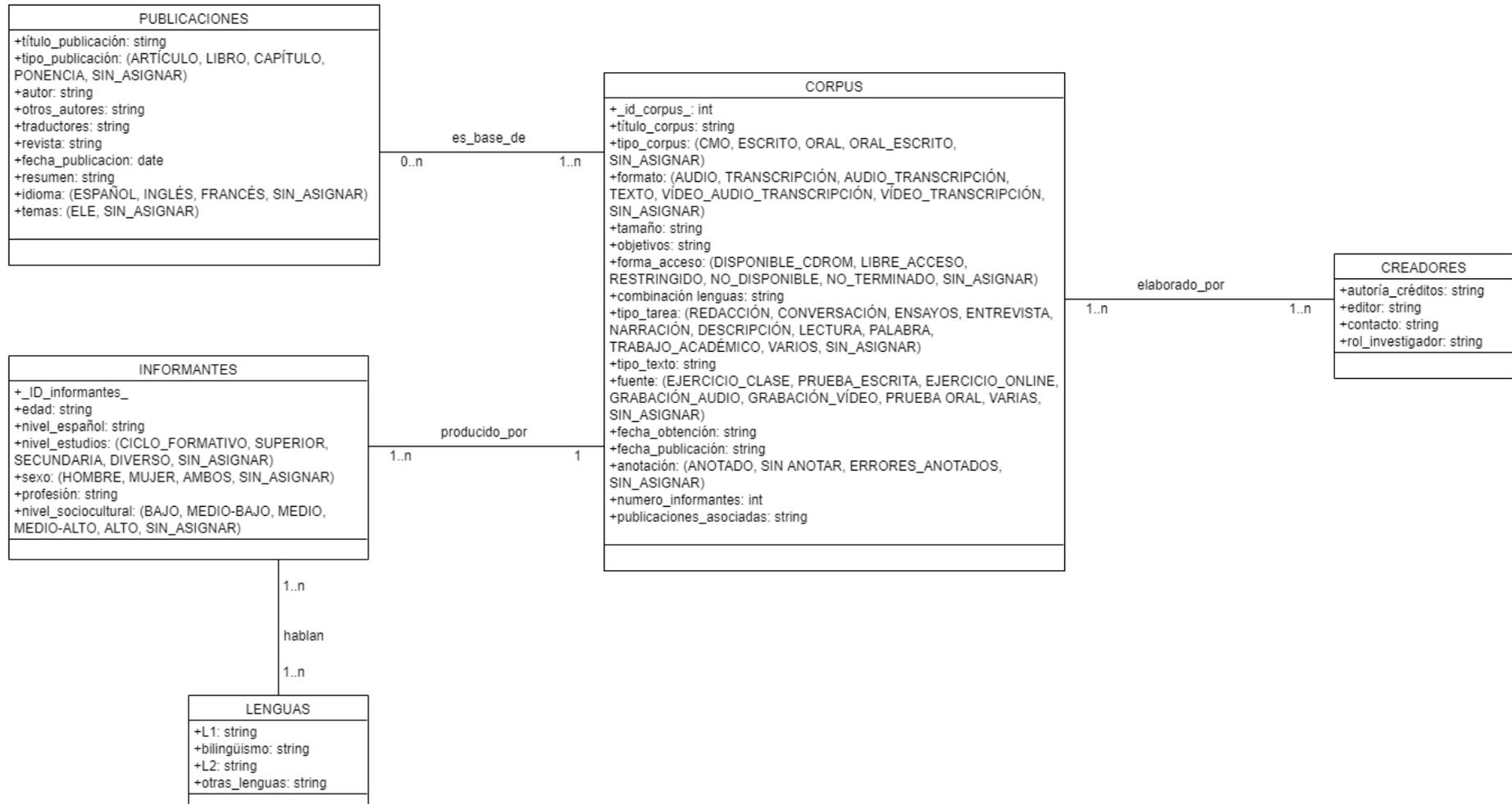
La seguridad estará también condicionada por las características de la aplicación OdA, de manera que habrá cuatro tipos de usuarios: superadministrador, administrador, usuario registrado y visitante ocasional. Las funciones de cada uno se detallan en la página 5 de la *Guía de gestión de OdA*. Los mecanismos de seguridad se implementarán por medio de un identificador de usuario (*login*) y una contraseña.

Anexo III: Modelo entidad-relación



En este modelo no están reflejados los atributos multivaluados ni modificaciones de última hora.

Anexo IV: diagrama UML



En este modelo no están reflejadas las últimas modificaciones que ha sufrido el indexador.

Anexo V: Cuestionario de evaluación empírica

INSTRUCCIONES

A continuación, encontrarás un cuestionario sobre el uso del Indexador de Corpus de Aprendices de Español. Está formado, primero, por una serie de casos de uso específicos del indexador que debes emular. Después, se te harán unas preguntas sobre aspectos generales.

Es importante que lo completes de forma individual y siguiendo las instrucciones que hay bajo estas líneas. Este cuestionario es anónimo y no tiene ningún tipo de implicación para ti, te rogamos que expreses tu opinión con total sinceridad.

1. Entra en el indexador (http://repositorios.fdi.ucm.es/corpus_aprendices_español).
2. Antes de empezar el cuestionario, dedica un momento a explorar la página principal y a familiarizarte con ella.
3. Lee la situación 1 y responde a las preguntas que aparecen en la tabla. Haz lo mismo con el resto de situaciones.
4. Dedicar cinco minutos a explorar libremente el indexador.
5. Responde a la segunda parte del cuestionario sobre aspectos generales.
6. Cuando hayas terminado, entrega el cuestionario completado.

No olvides anotar cualquier comentario o duda que te surja durante el cuestionario, luego tendremos tiempo para hablar sobre ello.

SITUACIÓN 1

Quieres encontrar todos los corpus con la siguiente característica:

- La lengua materna de todos o de varios informantes es el alemán

¿La búsqueda ha dado resultados?	SÍ <input type="checkbox"/>		NO <input type="checkbox"/>		
¿Has encontrado los resultados a través de la pestaña <i>búsqueda</i> o a través de la pestaña <i>navegación</i> ?	Búsqueda <input type="checkbox"/>		Navegación <input type="checkbox"/>		
Puntúa la facilidad de la búsqueda	Esta búsqueda me ha resultado....				
	Muy difícil <input type="checkbox"/>	Difícil <input type="checkbox"/>	Ni fácil ni difícil <input type="checkbox"/>	Fácil <input type="checkbox"/>	Muy fácil <input type="checkbox"/>
Describe y explica las dificultades que hayas encontrado (si no has encontrado dificultades no debes rellenar este cuadro)					

SITUACIÓN 2

Quieres encontrar todos los corpus que combinen las siguientes características:

- La lengua materna de los informantes es el francés
- Recogen muestras de lengua escrita
- Recogen textos narrativos

¿La búsqueda ha dado resultados?	SÍ <input type="checkbox"/>		NO <input type="checkbox"/>		
¿Has encontrado los resultados a través de la pestaña <i>búsqueda</i> o a través de la pestaña <i>navegación</i> ?	Búsqueda <input type="checkbox"/>		Navegación <input type="checkbox"/>		
Puntúa la facilidad de la búsqueda	Esta búsqueda me ha resultado....				
	Muy difícil <input type="checkbox"/>	Difícil <input type="checkbox"/>	Ni fácil ni difícil <input type="checkbox"/>	Fácil <input type="checkbox"/>	Muy fácil <input type="checkbox"/>
Describe y explica las dificultades que hayas encontrado (si no has encontrado dificultades no debes rellenar este cuadro)					

SITUACIÓN 3

Quieres encontrar todos los corpus con la siguiente característica:

- Son multilingües (es decir, recogen muestras de más de dos lenguas)

¿La búsqueda ha dado resultados?	SÍ <input type="checkbox"/>		NO <input type="checkbox"/>		
¿Has encontrado los resultados a través de la pestaña <i>búsqueda</i> o a través de la pestaña <i>navegación</i> ?	Búsqueda <input type="checkbox"/>		Navegación <input type="checkbox"/>		
Puntúa la facilidad de la búsqueda	Esta búsqueda me ha resultado....				
	Muy difícil <input type="checkbox"/>	Difícil <input type="checkbox"/>	Ni fácil ni difícil <input type="checkbox"/>	Fácil <input type="checkbox"/>	Muy fácil <input type="checkbox"/>

Describe y explica las dificultades que hayas encontrado (si no has encontrado dificultades no debes rellenar este cuadro)	
---	--

SITUACIÓN 4

Quieres encontrar todos los corpus que combinen las siguientes características:

- Recogen muestras de inglés y español
- Los informantes solo son mujeres

¿La búsqueda ha dado resultados?	SÍ <input type="checkbox"/>		NO <input type="checkbox"/>		
¿Has encontrado los resultados a través de la pestaña <i>búsqueda</i> o a través de la pestaña <i>navegación</i> ?	Búsqueda <input type="checkbox"/>		Navegación <input type="checkbox"/>		
Puntúa la facilidad de la búsqueda	Esta búsqueda me ha resultado....				
	Muy difícil <input type="checkbox"/>	Difícil <input type="checkbox"/>	Ni fácil ni difícil <input type="checkbox"/>	Fácil <input type="checkbox"/>	Muy fácil <input type="checkbox"/>
Describe y explica las dificultades que hayas encontrado (si no has encontrado dificultades no debes rellenar este cuadro)					

ASPECTOS GENERALES

Indica en qué medida estás de acuerdo con las siguientes afirmaciones:

(por favor, marca una sola casilla por afirmación)

	Totalmente en desacuerdo	En desacuerdo	Ni de acuerdo ni en desacuerdo	De acuerdo	Totalmente de acuerdo
1. Creo que usaría este indexador frecuentemente	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. Este indexador me parece innecesariamente complejo	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. Creo que el indexador ha sido fácil de usar	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. Creo que necesitaría ayuda de una persona con conocimientos técnicos para usar este indexador	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. Las funciones de este indexador están bien integradas	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. Creo que este indexador es muy inconsistente	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7. Imagino que la mayoría de la gente aprendería a usar este indexador rápidamente	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8. Encuentro que el indexador es muy incómodo de usar	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9. Me he sentido muy seguro usando el indexador	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10. He necesitado aprender muchas cosas antes de ser capaz de usar el indexador	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
11. Me parece adecuada la terminología utilizada en el indexador	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
12. Me parecen útiles todos los apartados que forman el indexador (navegación, búsqueda, colecciones, contacto y recomendaciones)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Anexo VI: Resultados de la evaluación empírica

Participant e	q1	q2	q3	q4	q5	q6	q7	q8	q9	q10	Puntuación SEU
1	3	5	2	5	3	4	2	4	2	3	27,5
2	2	2	4	1	3	3	2	2	4	3	60,0
3	1	3	3	4	3	4	2	4	2	4	30,0
4	3	3	1	4	3	3	3	4	2	3	37,5
5	3	3	4	2	3	3	5	3	3	1	65,0
6	3	3	3	4	3	3	2	3	2	2	45,0
7	2	4	2	4	4	3	2	4	1	3	32,5
8	3	4	2	5	4	2	4	3	2	4	42,5
9	4	2	1	4	2	5	2	1	1	3	37,5
10	3	4	3	4	3	3	3	3	3	3	45,0
11	1	1	3	4	4	3	5	3	4	1	62,5
12	1	4	4	4	4	3	4	4	3	4	42,5
13	3	4	3	3	2	4	2	3	4	4	40,0
14	4	2	4	1	4	2	3	2	4	2	75,0
15	4	3	1	2	4	3	2	3	2	3	47,5
16	4	2	3	2	3	2	4	2	2	3	62,5
17	3	2	2	4	3	3	4	1	3	3	55,0
18	3	2	2	4	3	3	4	1	3	3	55,0
19	3	2	4	1	4	2	4	2	4	1	77,5
20	3	4	2	4	2	3	3	3	2	2	40,0
21	3	4	2	5	3	3	4	3	1	3	37,5
22	4	3	2	1	4	3	3	2	3	2	62,5
23	3	2	2	4	3	2	5	1	3	2	62,5
24	2	4	1	5	2	3	2	4	1	3	22,5
25	2	4	2	4	2	4	1	5	2	3	22,5
26	3	3	2	4	3	3	3	4	2	4	37,5
27	5	1	3	4	3	2	5	1	3	3	70,0
28	4	2	2	4	4	3	4	3	2	3	52,5
29	4	3	2	4	4	3	4	3	3	3	52,5
30	3	4	2	2	3	3	5	3	2	2	52,5
31	3	2	3	2	3	3	2	3	4	2	57,5
32	3	4	3	5	3	3	1	3	2	1	40,0
33	3	4	2	5	3	1	4	4	2	3	42,5
34	3	4	3	5	3	2	4	4	1	3	40,0
35	2	4	2	4	3	3	4	2	2	3	42,5
36	4	3	4	4	4	3	4	2	4	4	60,0
37	2	3	2	3	4	2	3	3	2	4	45,0
38	1	3	2	5	5	3	2	4	1	4	30,0
39	3	4	2	4	3	2	3	2	3	4	45,0
PROMEDIO											47,6

Tabla 10: resultados del cuestionario de usabilidad SUS del grupo 1

Participante	q1	q2	q3	q4	q5	q6	q7	q8	q9	q10	SUS Score
1	1	5	1	5	3	3	1	5	1	5	10,0
2	3	4	3	2	2	4	4	3	2	4	42,5
3	3	4	1	5	3	3	2	4	1	5	22,5
4	3	5	1	4	3	3	2	4	1	4	25,0
5	2	4	2	5	4	3	4	4	1	3	35,0
6	4	3	4	1	4	3	4	2	3	1	72,5
7	3	4	3	1	2	4	4	3	3	2	52,5
8	4	4	4	3	4	3	2	3	3	2	55,0
9	3	3	3	2	3	3	1	3	2	2	47,5
10	3	3	3	2	2	3	2	3	2	1	50,0
11	3	3	2	1	4	3	5	2	3	2	65,0
12	1	4	2	4	3	3	1	4	1	4	22,5
13	4	4	2	5	2	3	3	4	2	3	35,0
14	3	2	3	3	4	2	3	2	3	2	62,5
15	3	4	2	4	2	3	2	3	2	3	35,0
16	4	4	4	4	3	3	4	3	3	4	50,0
17	2	4	3	3	3	4	3	2	2	2	45,0
18	2	3	2	4	3	3	2	4	2	4	32,5
19	3	3	1	5	2	3	5	4	3	4	37,5
20	1	3	1	4	1	3	1	5	1	5	12,5
21	4	4	3	2	4	2	2	2	3	3	57,5
22	3	4	2	3	2	3	2	3	2	3	37,5
23	4	4	1	4	3	3	2	4	2	4	32,5
24	4	3	2	3	4	2	4	3	3	3	57,5
25	4	2	4	3	3	2	3	2	3	2	65,0
26	4	2	5	1	4	1	5	5	3	1	77,5
27	2	4	2	4	3	3	2	4	2	4	30,0
28	2	4	1	5	3	4	2	4	1	4	20,0
29	3	4	2	3	3	4	3	3	2	3	40,0
30	3	4	2	4	3	4	3	3	2	4	35,0
31	4	2	3	2	4	1	4	1	3	2	75,0
32	4	2	3	2	3	2	4	3	3	1	67,5
33	3	4	2	4	3	3	2	4	3	2	40,0
34	3	4	3	4	4	3	2	3	2	3	42,5
35	3	4	2	4	4	3	2	4	2	3	37,5
36	4	2	4	2	4	3	4	2	4	4	67,5
PROMEDIO											44,2

Tabla 11: resultados del cuestionario de usabilidad SUS del grupo 2

Participante	resultados_B1	vía_B1	facilidad_B1	resultados_B2	vía_B2	facilidad_B2	resultados_B3	vía_B3	facilidad_B3	resultados_B4	vía_B4	facilidad_B4
1	si	n	2	no	NS/NC	NS/NC	no	NS/NC	NS/NC	si	n	2
2	si	n	3	no	b	3	si	n	2	no	b	2
3	si	n	3	no	b	3	si	n	2	no	b	2
4	si	n	3	no	b	3	no	b	3	no	b	1
5	si	b	5	si	b	3	no	b	1	no	b	1
6	si	n	4	si	n	4	si	n	4	no	n	2
7	si	b	3	no	b	1	no	b	1	no	b	1
8	si	b	4	no	b	2	no	NS/NC	2	no	b	2
9	si	n	2	no	NS/NC	NS/NC	no	NS/NC	NS/NC	si	n	2
10	si	b	3	no	b	3	si	b	3	no	b	3
11	si	b	4	si	b	4	no	b	3	no	b	3
12	si	n	2	no	b	4	si	n	5	no	b	4
13	si	n	2	no	b	4	si	n	4	no	b	3
14	si	n	4	si	n	3	si	n	3	si	b	2
15	si	b	2	no	b	3	no	b	2	no	b	2
16	si	n	3	no	b	2	no	NS/NC	2	no	n	2
17	si	b	4	si	b	3	si	b	3	si	b	3
18	si	b	4	si	b	3	si	b	2	no	n	1
19	si	n	3	si	NS/NC	4	no	n	2	si	n	5
20	no	b	2	no	b	2	no	b	1	no	b	2
21	si	b	2	no	b	4	no	b	2	no	b	NS/NC
22	si	b	4	no	NS/NC	NS/NC	no	NS/NC	NS/NC	NS/NC	b	NS/NC
23	si	b	5	si	b	4	no	b	2	no	b	1
24	si	n	1	no	NS/NC	NS/NC	si	n	4	no	b	1
25	no	NS/NC	2	no	NS/NC	1	si	n	5	no	n	3

26	si	b	2	no	NS/NC	1	no	b	2	no	b	1
27	si	b	5	no	b	3	no	b	1	no	b	1
28	si	n	3	NS/NC								
29	si	n	3	NS/NC								
30	si	b	5	no	b	3	no	NS/NC	NS/NC	si	b	2
31	si	b	3	no	NS/NC	3	si	n	3	no	b	3
32	si	b	3	no	n	2	si	n	3	si	n	4
33	si	b	4	no	b	4	no	b	4	no	b	NS/NC
34	si	b	4	no	b	NS/NC	no	b	NS/NC	no	b	NS/NC
35	si	b	2	no	b	2	no	b	2	no	b	2
36	si	b	3	si	b	3	si	b	3	si	b	3
37	si	b	1	si	b	3	si	b	1	si	b	1
38	si	b	2	no	NS/NC	1	si	NS/NC	NS/NC	no	b	2
39	si	b	3	si	b	1	si	b	4	si	NS/NC	2
40	si	n	3	no	b	2	no	b	2	NS/NC	NS/NC	NS/NC
41	si	b	3	si	b	3	si	b	3	si	b	4
42	si	NS/NC	2	no	NS/NC	1	si	NS/NC	2	no	NS/NC	1

Tabla 12: resultados a las situaciones del cuestionario (grupo 1)

Participante	resultados_B1	vía_B1	facilidad_B1	resultados_B2	vía_B2	facilidad_B2	resultados_B3	vía_B3	facilidad_B3	resultados_B4	vía_B4	facilidad_B4
1	no	NS/NC	1									
2	si	b	2	no	b	1	no	b	2	no	b	4
3	si	b	2	no	b	1	no	b	1	no	b	1
4	si	b	3	si	b	3	si	b	2	no	NS/NC	NS/NC
5	si	n	2	si	b	2	no	b	2	no	b	2
6	si	b	5	no	b	2	si	n	4	si	n	3
7	si	b	5	si	b	2	si	n	5	si	b	3
8	si	b	4	si	b	4	si	n	2	si	b	3
9	si	b	4	si	b	4	si	b	3	no	b	1
10	si	b	4	si	b	4	si	b	3	no	b	1
11	si	n	4	si	b	3	si	b	4	si	b	2
12	si	n	3	si	b	3	no	NS/NC	2	no	NS/NC	1
13	si	n	2	no	b	1	si	n	3	no	b	2
14	si	n	3	si	b	2	si	n	3	no	NS/NC	1
15	si	b	4	si	b	3	no	NS/NC	1	no	NS/NC	1
16	si	b	4	no	b	2	si	n	4	si	n	3
17	si	b	3	no	NS/NC	1	no	NS/NC	1	no	NS/NC	1
18	si	b	3	no	NS/NC	NS/NC	NS/NC	NS/NC	NS/NC	NS/NC	NS/NC	NS/NC
19	si	b	2	no	NS/NC	NS/NC	si	b	5	si	b	4
20	si	b	2	no	NS/NC	NS/NC	no	NS/NC	NS/NC	no	NS/NC	NS/NC
21	si	b	5	si	b	5	si	n	4	no	NS/NC	2
22	si	b	4	si	b	2	no	b	1	no	b	1
23	si	b	2	no		2	no	NS/NC	2	no	NS/NC	2
24	si	b	3	no	b	3	no	b	1	no	b	1
25	si	n	3	si	n	4	si	n	5	si	n	3

26	si	n	3	si	n	4	si	n	5	si	n	3
27	no	NS/NC	2	si	n	2	si	n	3	no	NS/NC	2
28	no	NS/NC	2	no	NS/NC	2	si	n	2	no	NS/NC	1
29	si	b	2	no	b	2	no	b	2	no	b	2
30	si	b	3	no	b	3	no	b	3	no	b	2
31	si	b	3	si	b	4	NS/NC	NS/NC	NS/NC	no	NS/NC	NS/NC
32	si	b	4	no	b	2	si	n	4	no	NS/NC	2
33	si	b	2	si	b	3	si	n	3	si	n	3
34	si	b	3	si	b	3	si	n	3	si	n	3
35	no	b	3	no	n	3	no	NS/NC	3	no	NS/NC	3
36	si	b	2	no	b	1	si	n	2	no	NS/NC	1
37	si	b	2	no	b	2	si	n	2	no	n	1
38	si	b	2	NS/NC								
39	si	b	4	no	b	NS/NC	si	n	4	si	n	4

Tabla 13: resultados a las situaciones del cuestionario (grupo 2)

Participante	TERMINOLOGÍA	UTILIDAD DE LOS APARTADOS
1	3	3
2	4	3
3	5	3
4	3	4
5	4	4
6	3	3
7	5	2
8	4	4
9	4	2
10	4	3

11	4	3
12	4	4
13	3	3
14	4	4
15	3	4
16	2	4
17	3	4
18	3	4
19	4	5
20	3	3
21	3	4
22	2	4
23	3	5
24	3	4
25	3	2
26	4	4
27	4	4
28	4	4
29	5	5
30	4	3
31	4	3
32	3	4
33	4	4
34	1	3
35	4	4
36	3	4
37	2	3
38	5	3

39	4	4
40	NS/NC	NS/NC
41	NS/NC	NS/NC
42	NS/NC	NS/NC

Tabla 14: resultados de las afirmaciones sobre la utilidad de la terminología y los apartados (grupo 1)

Participante	TERMINOLOGÍA	UTILIDAD DE LOS APARTADOS
1	3	3
2	5	3
3	4	4
4	3	1
5	3	4
6	5	5
7	2	3
8	4	4
9	4	4
10	4	2
11	4	4
12	4	4
13	3	4
14	4	4
15	4	4
16	4	4
17	3	3
18	5	3
19	3	3
20	3	2

21	5	2
22	3	4
23	3	3
24	4	4
25	4	4
26	4	4
27	4	3
28	3	2
29	3	4
30	3	3
31	3	5
32	3	4
33	3	3
34	4	3
35	4	4
36	NS/NC	NS/NC
37	NS/NC	NS/NC
38	NS/NC	NS/NC
39	4	5

Tabla 15: resultados de las afirmaciones sobre la utilidad de la terminología y los apartados (grupo 2)

Participante	Transcripción_B1	Categoría	Transcripción_B2	Categoría	Transcripción_B3	Categoría	Transcripción_B4	Categoría
1								
2								
3								
4								
5								
6								

7	Me he liado porque hay muchos campos posibles para buscar.	Formulario de búsqueda						
7		Navegación desapercibida						
8	Hay muchos campos para rellenar. Puede resultar confuso cuál es el que interesa.	Formulario de búsqueda	Ídem	Formulario de búsqueda	Ídem	Formulario de búsqueda	No estoy segura de cómo introducir los criterios de búsqueda en los campos.	Multivaluados
8		Navegación desapercibida		Navegación desapercibida		Navegación desapercibida		
9								

10	<p>La experiencia puede resultar un poco frustrante al solo poseer un modo avanzado de búsqueda. Sería idónea reducir las casillas de búsqueda en primera opción. Las opciones de búsqueda más recurrentes deberían ser más visibles que el resto.</p>	Formulario de búsqueda	<p>La guía de uso debe ser más extensa para resolver toda duda posible.</p>				<p>Añadir un apartado de dudas frecuentes</p>	
----	--	------------------------	---	--	--	--	---	--

10		Navegación desapercibida						
11			La única dificultad es que no estoy familiarizada con los filtros y las opciones que hay.	Desconocimiento del sistema	No sé dónde debería haber puesto las lenguas para que fuera multilingüe.	Navegación desapercibida	La complicación se da en seleccionar más de una lengua.	Multivaluados
12								
13								
14								
15			No sabía bien dónde poner los datos. Los he puesto donde creo que corresponde y no da resultado.	Navegación desapercibida	Ídem	Navegación desapercibida		
16								
17								
18								

19	He utilizado el menú de navegación y quizá no es muy intuitivo la <u>primera vez</u> .	Desconocimiento del sistema			El buscador no permite encontrar resultados y hay que usar el navegador .	Multivaluados		
20			"Tipo de tarea" es muy poco intuitivo para buscar un texto narrativo.	Confusión <i>tipo de tarea y tipo de texto</i>			He encontrado corpus sólo de informantes mujeres, pero no he sabido buscar muestras de inglés y español.	Multivaluados
21								
22								
23			Encontrar textos narrativos en "tipo de tarea" no me resulta	Confusión tipo de tarea y tipo de texto			Dificultad para buscar informantes que hablan	Multivaluados

			muy intuitivo.				inglés y español	
24								
25								
26								
27					He querido marcar multilingüe y he puesto "sí". No sé si he hecho bien	Multivaluados	No sé rellenar las casillas de las lenguas	Multivaluados
27						Navegación desapercibida		
28								
29								
30			No aparecen resultados que tengan muestra escrita	Confusión tipo de tarea y tipo de texto	Si intento buscar enfocándome en la lengua es imposible encontrar la información. Por el contrario,	Multivaluados		

					buscando por género sí me aparecen dos corpus, aunque no son en inglés.			
30						Desplegables preferidos		
30						Recomendaciones de uso desapercibidas		
31			Resultó confuso buscar un texto narrativo porque no sabía exactamente en qué sección estaba (tipo de tarea)	Confusión tipo de tarea y tipo de texto				
32			Hay demasiadas	Formulario de búsqueda	No hay una sección	Navegación desapercibida		

			secciones en la búsqueda y algunas no son claras.		específica para ello, no está clara.			
33								
34	El nivel de búsqueda es muy exhaustivo en comparación con la situación propuesta y eso lo hace confuso (tantas opciones para un único rasgo)	Formulario de búsqueda	Puede que no sepamos usar la herramienta	Desconocimiento del sistema			Si hubiera leído antes recomendaciones de uso lo hubiera sacado.	Recomendaciones de uso desapercibidas
35					Tendría que haber una opción de seleccionar si es o no multilingü	Navegación desapercibida		

					e, ya que no se sabe qué información poner en el espacio.			
36								
37								
38							Es difícil saber cómo insertar las dos lenguas en la parte "combinación de lenguas meta". (Ya lo he encontrado)	Recomendaciones de uso desapercibidas

Tabla 16: resultados de análisis cualitativo (grupo 1)

Participante	Transcripción_B1	Categoría	Transcripción_B2	Categoría	Transcripción_B3	Categoría	Transcripción_B4	Categoría
1	Complicado porque creo que la búsqueda debería de	Formulario de búsqueda	Las mismas que en la situación 2. Lo único es que aquí,	Formulario de búsqueda	Los mismos problemas que en la situación 1 y 2.	Formulario de búsqueda	Ídem	Formulario de búsqueda

	<p>ser más sencilla e intuitiva. Hay muchos campos para rellenar y solo con poner "alemán" en el título no me sale nada.</p>		<p>por lo menos, he conseguido poner "muestras de lengua escrita" y "textos narrativos" , pero no me ha salido nada.</p>					
2	<p>Debería explicar los diferentes formularios de una forma detallada, mediante una ayuda o manual de usuario.</p>	<p>Formulario de búsqueda</p>	<p>La búsqueda a través de los desplegados no es intuitiva, ya que requiere un conocimiento avanzado de cada una de las pantallas y en este caso de la</p>	<p>Desconocimiento del sistema</p>				

			búsqueda lo que requiere más tiempo.					
3					La opción de multilingüe debería ser un campo de elección y no para rellenar, porque no sé qué escribir.	Formulario de búsqueda	No sé rellenar los campos <i>inglés, español.</i>	Multivaluados
3						Navegación desapercibida		
4	Llegar a la forma de búsqueda y que al dar a intro no se realice la búsqueda.	Botón intro búsqueda			Que hay que escribir todas las posibles lenguas.	Navegación desapercibida		
5	He intentado hacerlo por búsqueda pero no	Formulario de búsqueda			Por navegación he visto los que salían en el	Formulario de búsqueda	No encuentro la manera de marcar inglés y	Multivaluados

	encontraba nada. Por navegación ha sido más sencillo en el apartado de informantes.				apartado multilingüe pero en búsqueda no sabía cómo ponerlo para marcar el criterio multilingüe .		español en la búsqueda. Con el criterio de mujeres no había problema.	
6			En "tipo de texto" no aparecen resultados predeterminados para elegir, así que he optado por seleccionar "Narración" en la pestaña de "tipo de tarea".	Formulario de búsqueda	Donde búsqueda es más complejo. Hay un apartado que pone "multilingüe", pero no está claro qué hay que poner en él.	Formulario de búsqueda	Ocurre lo mismo que en el anterior. En "Búsqueda" no está claro dónde o cómo buscar que las muestras sean en dos idiomas a la vez. El género si es fácil porque hay	Multivaluados

							una pestaña desplegable.	
6				Desplegables preferidos				Desplegables preferidos
6				Confusión tipo de tarea y tipo de texto				Recomendaciones de uso desapercibidas
7							Al introducir peticiones como "español e inglés" o "inglés y español" utilizando una conjunción no arroja resultados. A su vez si pones Inglés / Español en vez de Español / Inglés	Multivaluados

							tampoco sale nada.	
8					He intentado encontrar resultados a través de búsqueda y no lo he encontrado .	Navegación desapercibida	Solo da resultados si introduces "español / inglés". Si lo pones al revés no sale nada.	Multivaluados
8								
9					Solo he encontrado un corpus y porque en "multilingüe" he puesto "alemán".	Navegación desapercibida	La selección en la búsqueda resultado confusa.	Formulario de búsqueda
10					Encuentro solo uno y poniendo "alemán" en la casilla "multilingüe"	Navegación desapercibida		

					e" de la búsqueda.			
11								
12								
13	No están claras las opciones dentro de la búsqueda de los datos.	Formulario de búsqueda					No entiendo cómo se puede buscar unos idiomas concretos en la barra de búsqueda. Al escribir la lengua en combinación de lenguas meta no me da ningún resultado.	Recomendaciones de uso desapercibidas
13								Formulario de búsqueda

14			Método poco rápido y eficaz. Letra muy pequeña.	Letra pequeña	Tras conocer un poco la página web esta situación me resultó más fácil que las anteriores.			
15					A través de <i>búsqueda</i> no queda claro qué palabra o palabras debemos insertar en el cajetín <i>multilingüe</i> para que dé acceso a todos los idiomas disponibles . A través de la navegación no es posible consultar	Ventana taxonomía navegación	Por las mismas razones que en la situación 3.	

					los corpus multilingües de todas las lenguas a la vez, sino que ofrece opciones por pares (alemán-español, alemán-francés, etc.)			
						Formulario de búsqueda		
16								
17	El proceso es un poco complejo, las cosas que rellenar son muchas.	Formulario de búsqueda						
		Navegación desapercibida						
18					No sé qué tengo que escribir en el blanco	Formulario de búsqueda		

					de monolingüe para conseguir resultados.			
19								
20								
21								
22								
23							El sistema de búsqueda es muy complicado .	Formulario de búsqueda
24					No sé cómo ha de ponerse que es multilingüe , aparece un apartado con un campo a rellenar.	Formulario de búsqueda	Mismo problema. El género es fácilmente reconocible.	Recomendaciones de uso desapercibidas
24								Formulario de búsqueda
24								Desplegables preferidos

25	Al no estar familiarizada con la herramienta no sabía cómo empezar a buscar.	Desconocimiento del sistema						
26			Letra muy junta y pequeña.	Letra pequeña				
27								
28								
29					No me sale nada al poner "varios idiomas" en multilingüismo. No sé cómo buscarlo.	Formulario de búsqueda	No sé dónde situar los idiomas en la búsqueda.	Formulario de búsqueda
29						Navegación desapercibida		Recomendaciones de uso desapercibidas
30								
31	Solo al principio. Nada más acceder a	Formulario de búsqueda						

	la ventana de búsqueda parece más difícil.							
31		Navegación desapercibida						
32					No puede hacerse mediante búsqueda, debe hacerse mediante navegación , y este último no muestra claramente los idiomas que comprende .	Ventana pequeña izquierda	Búsqueda no es útil para buscar varios idiomas; cuando se quieren poner varios idiomas en bilingüismo o plurilingüismo, al introducir las lenguas buscadas no encuentra (no se especifica el formato con que escribirlo)	Multivaluados

32								Recomendaciones de uso desapercibidas
33	Hay muchas opciones	Formulario de búsqueda						
34	El buscador tiene muchas opciones que al principio hacen difícil la búsqueda.	Formulario de búsqueda						
35	No encontré dificultades pero considero que el diseño de la página debería ser intuitivo. Debería de estar mejor distribuido porque, en mi opinión, da	Formulario de búsqueda			Difícil porque tengo que visitar de casilla en casilla las combinaciones multilingües que me propone la página. No sé si he llegado a encontrar algún	Multivaluados		

	sensación de agobio.				resultado práctico.			
36								
37					Se tiene que ir una a uno con las opciones que salen.	Multivaluados		

Tabla 17: resultados del análisis cualitativo (grupo 2)