

**Nº REFERENCIA:**  
HUM2007-65966-CO2-02



**MINISTERIO DE CIENCIA E INNOVACIÓN**  
**DIRECCIÓN GENERAL DE PROGRAMAS Y TRANSFERENCIA DE CONOCIMIENTO**

**PROYECTOS I+D, ACCIONES ESTRATÉGICAS Y ERANETS**

**INFORME DE SEGUIMIENTO ANUAL**

<b>Investigador Principal:</b> Miren Igone Zabala Unzalu
<b>Título del Proyecto:</b> RICOTERM-3: Enriquecimiento de ontologías para la extracción de información de los discursos especializados en español y vasco
<b>Organismo:</b> Universidad del País Vasco
<b>Centro:</b> Facultad de Ciencias Sociales y de la Comunicación
<b>Departamento:</b> Filología Vasca (00113)
<b>Fecha de inicio del proyecto:</b> 01/10/2007
<b>Fecha de finalización del proyecto:</b> 30/9/2010

Fecha: 20 de marzo de 2009

**SR. SUBDIRECTOR GENERAL DE PROYECTOS DE INVESTIGACIÓN**  
**C/ Albacete 5, 28071 MADRID**

## A. ACTIVIDADES REALIZADAS Y GRADO DE CONSECUCIÓN DE LOS OBJETIVOS PROPUESTOS

1. Describa brevemente las actividades realizadas en el pasado año de desarrollo del proyecto. Indique si existe algún resultado a que haya dado lugar el proyecto durante ese periodo.

a) **Incorporación de nuevos investigadores:** Uno de los objetivos que nos proponíamos con este proyecto era el de actuar como aglutinadores de diferentes investigadores que trabajan de forma más o menos aislada sobre el discurso de especialidad y extracción de información de estos discursos.

- Con fecha de 9 de abril de 2008 se solicita la inclusión de seis nuevos investigadores en el proyecto: Antton Elozegi, Fc. Javier Alberdi, Axun Aierbe, Iñaki Ugarteburu, Julio Garcia y Patxi Angulo.
- El 19 de junio de 2008 se me comunica la aceptación de inclusión de los nuevos investigadores.
- En carta remitida el 2 de julio de 2008 al Subdirector General de Proyectos de Investigación, comunico la defunción de la investigadora Axun Aierbe.

b) **Reuniones de coordinación:**

- Durante el año 2008 realizamos dos **reuniones de coordinación entre las investigadoras principales** de los dos subproyectos coordinados que componen RICOTERM-3: una en el País Vasco (28/07/2008) y otra en Barcelona (03-05/11/2008) con la participación del investigador Joseba Ezeiza.
- También se han realizado dos reuniones de coordinación que han reunido a todos los investigadores del subproyecto del País Vasco: 25/02/2008 y 04/02/2009.

c) **Convenios con organismos y empresas ajenas a la Universidad del País Vasco:**

- Se han realizado varias reuniones con empresas y organismos que trabajan en el procesamiento automático del lenguaje y extracción de información de discursos especializados: Ametzagaiña, Fundación Elhuyar, UZEI.
- Se ha materializado y firmado un convenio marco de colaboración entre la UPV y la empresa Ametzagaiña <<http://www.ametza.com/>>, empresa con la que estamos ya trabajando en la elaboración de corpus y extracción de información.
- Hemos redactado y pactado las condiciones de un convenio marco de colaboración con la Fundación Elhuyar <<http://www.elhuyar.es/>> y estamos llevando a cabo los trámites para que sea firmado cuanto antes este convenio. Mientras tanto Elhuyar ofrece condiciones privilegiadas de utilización de sus recursos textuales (corpus ZT) y sus extractores de terminología (Erauzterm y ELEXBI) para ayudar a la consecución de los objetivos de este proyecto. Por otro lado, esta fundación está trabajando en el enriquecimiento de EuroWordNet para el vasco en campos especializados, y hemos detectado una importante complementariedad con el trabajo que estamos realizando en RICOTERM-3.
- Hemos tenido un primer contacto con UZEI (Centro Vasco de Terminología y Lexicografía) <<http://www.uzei.com/home.asp>>. Este organismo se ha mostrado muy interesado en los objetivos del proyecto RICOTERM-3, y estamos trabajando en las modalidades y convenio de colaboración que sería adecuado para dar un grado adecuado de organización y sistemacidad a los diferentes modos de colaboración.

d) **Seminarios con otros grupos que trabajan sobre terminología, extracción y enriquecimiento de ontologías:**

- Varios investigadores de RICOTERM-3 hemos participado en seminarios con investigadores que trabajan en extracción de terminología y enriquecimiento de

EuroWordNet en el País Vasco para compartir información sobre nuestros proyectos y buscar líneas de colaboración y complementariedad: grupo de investigación IXA <<http://ixa.si.ehu.es/ixa> > y Fundación Elhuyar <<http://www.elhuyar.es/>>.

- a) Hondarribia (junio 2008): Grupo IXA e investigadores de RICOTERM-3
- b) Donostia (julio 2008): Grupo IXA, investigadores de Elhuyar e investigadores de RICOTERM-3
- c) Usurbil (noviembre 2008): Grupo IXA, investigadores de Elhuyar e investigadores de RICOTERM-3

**e) Elaboración de corpus:**

- Hemos trabajado en la localización de textos en euskera y hemos comenzado a elaborar corpus para la extracción de terminología que nos permita el enriquecimiento de EWN en tres campos especializados: Derecho, Medio Ambiente e Informática. < <http://weblingua.ametza.com/corpus-KL/> >
- Estamos trabajando en un protocolo para el procesamiento de los textos que tenga en cuenta la evaluación de varios extractores de terminología que utilizan diferentes etiquetados y/o diferentes cadenas de procesamiento: es necesario reservar parte de los corpus sin procesar para la evaluación y utilizar formatos estándar (TEI) para que el formato no de lugar a dificultades en la evaluación.

**f) Resultados del proyecto:**

Aunque en el proyecto no se preveía plan de formación ni becarios, uno de los objetivos era conseguir reunir un equipo de trabajo interdisciplinar y joven que reuniera sus esfuerzos en torno a la investigación de los discursos de especialidad. Uno de los investigadores (**Joseba Ezeiza**) ha realizado dos **cursos de postgrado** con sus correspondientes **memorias finales**. Estos trabajos son una aportación muy importante a los objetivos del proyecto.

EZEIZA, J. (2008c). "Especialitate hizkeren lantegi birtuala" [Taller Virtual de Lenguas de Especialidad]. Análisis preliminar para el desarrollo de una infraestructura telemática para la investigación y didáctica de la terminología en la universidad. Memoria de Diploma de Postgrado. Bartzelona: IULA

Publicación electrónica:

<URL: <http://brangaene.upf.es/moodle/login/index.php>>

EZEIZA, J. (2009). *PROYECTO GARALEX. La terminología en las producciones académicas del área de las Ciencias Jurídicas. Aproximación didáctica. Masterraren amaierako memoria*. Bartzelona:

IULA . Publicación electrónica:

<URL: <http://brangaene.upf.es/moodle/login/index.php>>

Por otra parte, la investigadora **Maria Jesus Aranzabe** ha presentado su tesis doctoral codirigida por el Dr. Jose María Arriola también investigador de este proyecto. Esta tesis es una importante aportación al proyecto RICOTERM-3 ya que nos aporta una herramienta muy valiosa para la extracción (semi)automática de información sintáctica de los corpus. Con este recurso nos situamos en situación comparable al grupo de la UPF, que cuenta ya con un banco de árboles sintácticos semejante para el español.

M<sup>a</sup> JESUS ARANZABE URRUZOLA (2008) *Dependentzia.ereduan oinarritutako baliabide sintaktikoak: zuhaitz bankua eta gramatika konputazionala* [Recursos sintácticos basados en

el modelo de dependencia: banco de árboles sintácticos y gramática computacional]  
DIRECTORES: JOSE MARIA ARRIOLA & MARIA ARANZAZU DIAZ DE ILLARRAZA  
Tesis doctoral de la UPV/EHU

**g) Contratación de técnicos.**

Hemos decidido retrasar la contratación de técnicos informáticos para el momento en que tuviéramos perfectamente definido el trabajo a realizar por dichos técnicos. Por este motivo no se han contratado aún. Por el momento hemos definido el trabajo a realizar por dichos técnicos y hemos definido su perfil para solicitar de inmediato su contratación. La idea es que dichos técnicos empiecen a trabajar antes del verano en la adecuación de los formatos de los corpus, extracción de información lingüística para facilitar el trabajo de los lingüistas y enriquecimiento de EWN.

2. Si ha encontrado problemas en el desarrollo del proyecto, coméntelos, especificando su naturaleza (de carácter científico, de gestión, etc).

A lo largo del desarrollo del proyecto nos hemos encontrado con dos **problemas** en cuanto al personal investigador:

- a) **Defunción** de la investigadora **Axun Aierbe** en 2008.
- b) **Enfermedad grave** del investigador **Iñaki Ugarteburu**. Se prevé que la baja se prolongue al menos durante seis meses más.

## **B. PERSONAL ACTIVO EN EL PROYECTO DURANTE EL PERÍODO QUE SE JUSTIFICA.**

En el cuadro siguiente debe recogerse la situación de todo el personal del o de los Organismos participantes que haya prestado servicio en el proyecto en la anualidad que se justifica, o que no haya sido declarado anteriormente, y cuyos costes (salariales, dietas, desplazamientos, etc.), se imputen al mismo.

Si la persona estaba incluida en la solicitud original, marque “S” en la casilla correspondiente y no rellene el resto de casillas a la derecha.

Indique en la casilla “Categoría Profesional” el puesto de trabajo ocupado, el tipo de contratación: indefinida, temporal, becarios (con indicación del tipo de beca: FPI, FPU, etc.), etc.

En el campo “Función en el proyecto” indique el tipo de función/actividad realizada en el proyecto, (p.ej., investigador, técnico de apoyo,...).

### **Recuerde que:**

- En este capítulo sólo debe incluir al personal vinculado a los Organismos participantes en el proyecto. Los gastos de personal externo (colaboradores científicos, autónomos...) que haya realizado tareas para el proyecto debe ser incluido en el capítulo de “Varios”.

- Las “Altas” y “Bajas” deben tramitarse de acuerdo con las “Instrucciones para el desarrollo de los proyectos de I+D” expuestas en la página web del MICINN.

Apellido 1	Apellido 2	Nombre	NIF/NIE	Catgª Profesional	Incluido en solicitud original	Si no incluido en solicitud original:		
						Función en el proyecto	Fecha de Alta	Observaciones
Zabala	Unzalu	Miren Igone	30551522 R	Investigadora principal	S			
Lersundi	Ayestaran	Mikel	34107733K	Investigador	S			
Aranzabe	Urruzola	Mª Jesus	72437201N	Investigadora	S			
Arriola	Egurrola	Jose Mª	78866445C	Investigador	S			
Elorduy	Urkiza	Agurtzane	11906545C	Investigadora	S			
San Martin	Eguia	Iciar	15389693W	Investigadora	S			
Elosegi	Aduriz	Mª Cristina	15874534A	Investigadora	S			
Ezeiza	Ramos	Joseba	15986557Q	Investigador	S			
Martínez	Bilbao	Mercedes	14581255Z	Investigadora	S			
Elosegi	Aldasoro	Antton	15221514E	Investigador		Elaboración de corpus de Derecho y extracción de terminología	02 /07 / 2008	
Alberdi	Larizgoitia	Fco.	14957811Z	Investigador		Elaboración de corpus e	02 /07 / 2008	

		Javier				investigación de estrategias lingüísticas		
García	García de los Salmones	Julio	14581066D	Investigador		Elaboración de corpus e investigación de estrategias lingüísticas	02 /07 / 2008	
Angulo	Martin	Patxi	15947604W	Investigador		Elaboración de corpus e investigación de estrategias lingüísticas	02 /07 / 2008	
Ugarteburu	Gastañares	Iñaki	72564427W	Investigador		Elaboración de corpus e investigación de estrategias lingüísticas	02 /07 / 2008	De baja por enfermedad grave.
Aierbe	Mendizabal	Axun	72437014D	Investigadora			02 /07 / 2008	Fallecida

**C. PROYECTOS COORDINADOS** (*Cumplimentar sólo por el coordinador si se trata de un proyecto coordinado*)

Describe el desarrollo de la coordinación entre subproyectos en este año, y los resultados de dicha coordinación con relación a los objetivos globales del proyecto.

**D. RELACIONES O COLABORACIONES CON DIVERSOS SECTORES**

1. En caso de que estuviera prevista la participación o respaldo activo por parte de alguna Empresa o Agente socio-económico con interés en el proyecto, indique cómo se está realizando dicha participación.

2. Si el proyecto ha dado lugar a otras colaboraciones o transferencias con entidades no académicas, descríbalas brevemente.

- **Convenio marco de colaboración con la empresa Ametzagaiña** apoyo tecnológico para la elaboración de corpus y extracción de información lingüística.
- **Colaboración con la Fundación Elhuyar:** compartir corpus y herramientas para la extracción de terminología. Complementariedad en el enriquecimiento de EWN para campos especializados en euskera.
- **Colaboración con UZEI:** compartir glosarios terminológicos y recursos textuales.

3. Si el proyecto ha dado lugar a colaboraciones con otros grupos de investigación, coméntelas brevemente.

- **Colaboración estrecha con el grupo de investigación IXA:** Tres de los investigadores de este proyecto (Mikel Lersundi, Jose M<sup>a</sup> Arriola y Maria Jesús Aranzabe) son también miembros del grupo de investigación IXA. Dado que este grupo participa en otros proyectos relacionados con textos especializados, terminología y enriquecimiento de EWN, la colaboración con este grupo es en todo momento muy estrecha.

4. Si ha iniciado la participación en proyectos del Programa Marco de I+D de la UE y/o en otros programas internacionales en temáticas relacionadas con la de este proyecto, indique programa, tipo de participación y beneficios para el proyecto.

**E. GASTOS REALIZADOS**

**Nota:** Debe cumplimentarse este apartado independientemente de la justificación económica enviada por el organismo.

1. Indique el total de gasto realizado en el proyecto hasta este momento:

	Total gasto de la anualidad (€)
Personal	0,00 €
Otros costes de ejecución	2536,47 €
<b>TOTAL GASTO REALIZADO</b>	<b>2536,47 €</b>

2. Comente brevemente si ha habido algún tipo de incidencia en este apartado.

**No hay incidencias reseñables.**

## **F. GASTOS DE EJECUCIÓN: MODIFICACIONES DE CONCEPTOS DE GASTO CON RESPECTO A LA SOLICITUD ORIGINAL.**

Recuerde que los trasvases entre gastos de personal y gastos de ejecución deben tramitarse de acuerdo con las “Instrucciones para el desarrollo de los proyectos de I+D” expuestas en la página web del MICINN.

a) **Equipamiento:** 808,37 €

En el cuadro adjunto, rellene una línea por **cada equipo adquirido no previsto en la solicitud inicial** que dio lugar a la concesión de la ayuda para el proyecto y justifique brevemente su adquisición. Si se ha adquirido un equipo en sustitución de otro que figuraba en la solicitud de ayuda inicial (por mejorar sus prestaciones, por obsolescencia del anterior...), indicarlo también en la casilla correspondiente.

<b>Identificación del equipo</b>	<b>Importe</b>	<b>Justificación adquisición</b>	<b>Sustituye a ...(en su caso).</b>

b) **Viajes/Dietas:** 1421,6 €

En el cuadro adjunto se justificará la imputación de gasto en viajes y dietas sólo en el caso de que este **tipo de gasto no estuviera previsto en la solicitud inicial.**

--



**c) Material fungible: 279,6 €**

Se describirá y razonará en el siguiente cuadro la adquisición del material fungible incluido en la justificación, sólo cuando **este tipo de gasto no estuviera previsto en la solicitud original**.

--

**d) Varios: 26,9**

Se describirán en el siguiente cuadro los gastos varios más relevantes incluidos en la justificación y **no previstos en la solicitud original**, justificando brevemente su inclusión. En este apartado se incluirá, entre otros, al personal externo y, en el caso de que el gasto justificado se refiera a colaboraciones científicas, se identificará al colaborador.

--