

BASYQUE: Aplicación para el estudio de la variación sintáctica

Larraitz Uriia
Centro IKER (UMR5478) – IKERBASQUE
Baiona-Bayona (Francia)
larraitz.uria@ehu.es

Ricardo Etxepare
Centro IKER (UMR5478)
Baiona-Bayona (Francia)

Resumen

En este artículo presentamos BASYQUE, la aplicación que hemos desarrollado para el estudio de la variación sintáctica. Aunque este proyecto se centra fundamentalmente en los dialectos del País Vasco francés (*Iparralde*), BASYQUE es una aplicación multilingüe útil también para el análisis de otras lenguas y/o dialectos. Además de presentar las opciones que ofrece BASYQUE, abordaremos los aspectos metodológicos y los criterios que hemos seguido para realizar esta aplicación. El desarrollo de este proyecto cuenta con la colaboración del Centro IKER con sede en Bayona (Francia) y el Grupo IXA de la Universidad del País Vasco (UPV).

1. Introducción

Durante los últimos años ha ido cambiado considerablemente el punto de vista analítico sobre el uso del código estándar de una lengua y sus dialectos: en contra de la tesis que defendía un único modelo lingüístico, basado en el uso exclusivo y generalizado del código estándar, en la actualidad cobra fuerza la idea de que el código estándar no garantiza una comunicación eficaz en todos los contextos o situaciones comunicativas y que, por lo tanto, conviene abrir un espacio de uso también para los códigos comunicativos propios de cada lugar, es decir, para las variantes dialectales territoriales.

Desde el reciente interés suscitado por los registros dialectales en la constitución de corpus lingüísticos y con la estimable ayuda de los recursos que nos ofrecen las nuevas tecnologías en el campo del Procesamiento del Lenguaje Natural (PLN), hemos desarrollado la aplicación denominada BASYQUE¹. Esta aplicación nos brinda la posibilidad de realizar diferentes tipos de consulta y obtener así información precisa sobre estructuras sintácticas concretas que presentan alternancia interdialectal. Se trata, por tanto, de un soporte que podemos definir como un atlas sintáctico básico.

La lengua vasca (*euskara*) constituye en la actualidad un campo de investigación privilegiado para el estudio de la variación sintáctica, ya que presenta un entorno lingüístico muy rico en la que

aún concurren y conviven diversas variantes dialectales.

En este proyecto nos centramos en el análisis de la variación sintáctica que existe entre los dialectos del País Vasco francés (*Iparralde*). Precisamente en *Iparralde*, el carácter dialectal de los textos escritos es, con diferencia, más acusado que en cualquier otra zona del País Vasco. Ello se debe, entre otros factores, a que a pesar de que el código estándar (*euskara batua*) viene incrementando su presencia progresivamente, éste se introdujo más tardíamente en un contexto que además no conoce, de momento, una escolarización masiva en la lengua vasca. En este contexto, las variantes dialectales siguen teniendo un vigor considerable, tanto en el ámbito de la literatura escrita como en los distintos medios de comunicación que operan en *Iparralde*.

Por otra parte, la variación sintáctica presente en los dialectos de *Iparralde* no ha sido estudiada detenidamente desde el punto de vista gramatical, por lo que no disponemos ni de textos lingüísticamente etiquetados ni bases de datos, infraestructuras o entornos informáticos que proporcionen información detallada sobre las características de tales dialectos.

Consideramos, por tanto, que el desarrollo de herramientas que faciliten el análisis de estas variantes dialectales constituye una aportación interesante desde el punto de vista de los corpus disponibles en soporte informático y contribuye a una representación más adecuada de la riqueza lingüística existente en el ámbito del *euskara*. Con

¹ <http://ixa2.si.ehu.es/atlas2/index.php?lang=es>

ese objetivo hemos desarrollado BASYQUE. Y aunque este estudio se centra en los dialectos del País Vasco francés, BASYQUE es una aplicación multilingüe útil también para el tratamiento y análisis de otras lenguas y/o dialectos.

Este proyecto viene desarrollándose con la colaboración de dos grupos de investigación de amplia experiencia en este campo: el centro IKER (UMR5478) de Bayona (Francia), que trabaja en el ámbito de la lingüística y la dialectología de la lengua vasca, y el grupo IXA de la Universidad del País Vasco (UPV), que lleva años trabajando en la creación de recursos informáticos básicos para el PLN en general y para el análisis del *euskara*, en particular².

Tras esta breve introducción, nos centraremos primeramente en la metodología que hemos aplicado y en los criterios que hemos definido para la recogida de datos (apartado 2). En el apartado 3, presentaremos BASYQUE y daremos a conocer el tipo de información que puede consultarse en esta aplicación. Finalmente, en el apartado 4, trataremos de exponer unas primeras conclusiones y abordaremos las líneas futuras de trabajo.

2. Metodología y criterios para la recogida de datos

Para poder efectuar consultas en BASYQUE y obtener así información en torno a la variación sintáctica, es imprescindible alimentar primero la base de datos. Para ello, hay determinados aspectos metodológicos que hay que tener en cuenta en este tipo de atlas sintácticos: la delimitación del campo de trabajo, la selección de informantes, la elección de las variables lingüísticas que serán objeto de análisis, la elaboración de cuestionarios y, por último, la recogida de datos.

² El origen de este trabajo se encuentra en el proyecto TSABL (*Towards a Syntactic Atlas of the Basque Language*: <http://www.iker.cnrs.fr/-tsabl-towards-a-syntactic-atlas-of-.html?lang=fr>) y esta ligado al proyecto BasDiSyn (*Basque Dialect Syntax*: <http://basdisyn.net/en/>). BasDiSyn es un grupo compuesto por investigadores que trabajan en el estudio de la micro-variación sintáctica del *euskara*. Este grupo, a su vez, se encuentra en la red europea de investigación Edysin (<http://www.dialectsyntax.org/index.php/home-mainmenu-1>), en colaboración con otros investigadores que también se dedican a investigar en el campo de la dialectología (Carrilho, 2010; Barbiere and Bennis, 2007; entre otros).

En lo que respecta a la delimitación del campo de trabajo, nuestro proyecto abarca los dialectos del País Vasco francés. *Iparalde* cuenta con una comunidad lingüística vasca relativamente modesta de unos 55.000 hablantes (para un total de 150.000 habitantes, aproximadamente) y donde el proceso de estandarización de la lengua vasca avanza a un ritmo más lento que en el resto del País Vasco. Así, el carácter dialectal de los textos escritos en dicho territorio es más marcado que en cualquier otra zona del país, por lo que nos encontramos en un entorno que podemos calificar de excepcional para el estudio de la variación sintáctica.

Para llevar a cabo este proyecto hemos seleccionado 20 localidades, tomando en cuenta la subdivisión dialectal del País Vasco francés.

En cada una de las localidades, seleccionamos cuatro informantes que deben de cumplir los siguientes requisitos:

- Tener como lengua materna el dialecto de su localidad.
- Haber nacido, tanto el informante como sus padres, en esa misma localidad.
- No haber residido durante un período superior a siete años fuera de la localidad natal.
- Tratarse de informantes hablantes activos de su dialecto.

De estos cuatro informantes, hemos de contar con la colaboración de una persona mayor de entre 50 y 70 años, otra de entre 36-50 años y dos jóvenes comprendidos en la franja de 18-35 años, de los cuales uno realiza o ha realizado sus estudios en *euskara* y el segundo no. Este último perfil busca la posibilidad de analizar i) hasta qué punto las formas dialectales locales siguen vigentes entre los jóvenes, ii) la incidencia potencial de los procesos de escolarización en el uso/abandono de las variantes dialectales, y iii) la influencia que puede ejercer en la juventud el contacto con una lengua dominante como es el francés.

Somos conscientes de que este criterio de selección tan riguroso puede dificultar la búsqueda de informantes apropiados, pero resulta un factor imprescindible para que los datos recopilados ofrezcan una visión lo más uniforme posible y para que las conclusiones sean debidamente contrastadas.

La elección de los fenómenos lingüísticos a analizar depende fundamentalmente de las características de

cada lengua y/o dialecto. Las alternancias sintácticas en las que nos centramos en este proyecto están ligadas a la concordancia, la determinación, las estructuras posposicionales, la nominalización, las expresiones de modalidad y la evidencialidad. Todos ellos son aspectos centrales de la gramática del *euskara* en los que tiende a producirse alternancia o variación sintáctica interdialectal. A modo de ejemplo, presentamos algunas estructuras donde ocurre variación sintáctica:

- Concordancia entre el argumento dativo y el verbo:

Una regla gramatical de aplicación general en la lengua vasca exige que el argumento dativo concuerde en número y persona con el auxiliar. Esta regla general, que no conoce excepciones en el habla del País Vasco peninsular, tiene un estatus opcional en el *euskara* de *Iparralde*. Por tanto, es posible encontrarse con alternancias del tipo representado en (1a, b):

(1) a. *Ez dio neori deus igorri*
 Ez-NEG dio-AUX neori-DAT deus-ABS igorri-V
 abs/erg/dat
 No ha enviado nada a nadie

b. *Ez du neori deus bota*
 Ez-NEG du-AUX neori-DAT deus-ABS bota-V
 abs/erg
 No ha enviado nada a nadie

Esta opcionalidad está sujeta a varios condicionantes, que tienen que ver con la naturaleza referencial del argumento dativo y las propiedades léxicas del verbo.

- Posición de la partícula evidencial *omen* (“según parece”, “por lo visto”):

En algunas localidades la partícula *omen* aparece entre el verbo y el auxiliar (2a), mientras que en otras zonas el auxiliar se antepone al verbo principal y *omen* se coloca entre el auxiliar y el verbo (2b):

(2) a. *Langonen bizi omen zen*
 Langonen-INE bizi-V omen-PRT zen-AUX
 Según parece residía en Langon

(2) b. *Langonen zen omen bizi*
 Langonen-INE zen-AUX omen-PRT bizi-V
 Según parece residía en Langon

Estas diferencias parecen estar relacionadas con la categoría léxica de la partícula evidencial: en el primer caso, ésta se comporta como una cabeza funcional; en el segundo caso, como un adverbio.

- Alternancia en el uso del determinante:

En algunas zonas de *Iparralde* se omite el determinante del sustantivo en posición de objeto y el sustantivo tiene una lectura plural (3a), mientras que en otras localidades tal omisión es imposible (3b):

(3) a. *Hor bada sagar*
 Hor-ADV bada-V sagar-N
 Ahí hay manzanas

(3) b. *Hor badira sagarrak*
 Hor-ADV badira-V sagar-N ak-DET
 Ahí hay manzanas

Ejemplos de alternancia sintáctica como los arriba indicados son precisamente los que pretendemos recopilar y clasificar en BASYQUE. Así, los cuestionarios empleados indagan sobre la distribución geográfica de estas variantes.

Para la recopilación de los datos que nos interesan, disponemos de tres fuentes de información. La fuente principal de información la constituyen los cuestionarios. Cada cuestionario aborda un fenómeno lingüístico concreto y son cuestionarios específicos diseñados para elicitar ejemplos concretos de variación sintáctica interdialectal. Preparamos diferentes tipos de tests, que consisten principalmente en frases para traducir y tareas de selección de variantes basadas en una escala de aceptabilidad relativa. En algunos casos, pedimos al encuestado que complete una forma lingüística.

Pasamos los cuestionarios a los informantes previamente seleccionados, grabamos las respuestas y guardamos los resultados obtenidos (tanto las respuestas como la información correspondiente a cada una de ellas) en una base de datos a la que podemos acceder mediante la aplicación BASYQUE.

Aparte de las encuestas diseñadas específicamente para BASYQUE, existe en el País Vasco una amplia tradición investigadora en el campo de la dialectología, en la que abundan ejemplos característicos y significativos de las variantes dialectales del *euskara*. Además, en los textos literarios también están presentes las múltiples

alternancias sintácticas. Por tanto, y aunque los cuestionarios constituyan la principal fuente de información para la materialización de este estudio, hemos considerado interesante reunir en una misma aplicación ejemplos de origen diverso. Ello nos proporciona, por una parte, mayor abundancia de ejemplos y, por otra, la posibilidad de establecer comparaciones entre los datos obtenidos mediante cuestionarios específicos con los ejemplos obtenidos por medio de otras fuentes de información. De hecho, recopilar ejemplos de origen diverso es importante debido a que ninguna de las fuentes mencionadas tiene un carácter óptimo para el estudio de la variación sintáctica, es decir, todas presentan algunos pros y contras (véase, por ejemplo, Cornips, L. & C. Poletto, 2005). Obviamente, la base de datos consigna el origen distinto de cada una de las fuentes.

El proceso que conduce desde la recopilación de datos hasta su consulta, puede resumirse de la manera siguiente:

- Preparación previa de cuestionarios específicos que nos permitan obtener ejemplos de fenómenos lingüísticos concretos y búsqueda de ejemplos de dichas variantes dialectales tanto en grabaciones realizadas en otros proyectos como en textos literarios recientes.
- Introducción de los ejemplos obtenidos en una base de datos, junto con la correspondiente información adicional³.
- Consulta sistemática de todos los datos almacenados con la ayuda de la aplicación BASYQUE.
- Análisis de datos y conclusiones.

Gracias a la información recopilada en BASYQUE, contamos con una base muy sólida (y en este caso única) que nos permite continuar progresando en el estudio de la variación sintáctica dialectal de *Iparralde*, tanto desde el punto de vista lingüístico como en el campo del PLN.

3. La aplicación BASYQUE

³ Como detallaremos más adelante, junto a cada ejemplo, especificamos la escritura formal de la forma dialectal transcrita, el texto normalizado, la información gramatical de cada elemento de la frase (glosas), las propiedades lingüísticas que le corresponden, el cuestionario al que pertenece, localidad en la que se ha recogido el ejemplo, etc.

BASYQUE es la aplicación que hemos desarrollado para el estudio de la variación sintáctica que se produce entre dialectos.

En la imagen 1 (última página) puede verse la interfaz de la aplicación, donde se encuentran las principales opciones disponibles. En el apartado *Información* se recogen las aportaciones relacionadas con este trabajo o proyectos similares. La sección *Administración* está destinada a aquellos usuarios que se registran por primera vez (el registro no es condición necesaria para poder realizar consultas). En *Contacto* se especifican nuestros datos para que los usuarios puedan contactarnos para aclarar dudas, resolver problemas o realizar nuevas propuestas. En el apartado *Ayuda* el usuario dispone de un manual en el que se describen detalladamente las formas de uso y las posibilidades que ofrece la aplicación. Junto al apartado de *Ayuda*, se puede seleccionar el *idioma* en el que se quiere realizar las consultas: euskara, español, inglés o francés. En la parte superior de la interfaz se localiza el *Campo de Registro* desde el cual los anotadores y el administrador acceden a la sección privada de la aplicación para así introducir, modificar o eliminar datos. En la parte derecha se encuentra el *Menú de búsqueda* donde están disponibles diversas opciones de consulta. Las respuestas obtenidas en cada búsqueda se visualizan en el *Mapa* que se encuentra localizado en el centro mismo de la interfaz.

En lo que se refiere a la búsqueda de datos, BASYQUE ofrece la posibilidad de realizar diferentes tipos de consulta y de buscar ejemplos que provienen de tres fuentes de información diferentes: cuestionarios, conversaciones libres grabadas en otros proyectos y corpus literarios. Cada fuente de información cuenta con una serie de opciones y atributos de búsqueda que detallamos a continuación.

3.1. Consulta de ejemplos obtenidos mediante cuestionarios

Los ejemplos obtenidos mediante cuestionarios específicos constituyen la fuente de información principal de este proyecto. Las respuestas obtenidas pueden consultarse a distintos niveles: por localidades, por meta-categorías lingüísticas o por clases de verbos, por cuestionarios, por propiedades lingüísticas y también por edades de los informantes.

Al seleccionar una de las opciones mencionadas, en el menú de búsqueda se muestran los atributos disponibles que le corresponden a cada una de ellas. Por ejemplo, si seleccionamos la opción *Localidad*, se visualizan las localidades en las que se han realizado los cuestionarios o si optamos por las *Propiedades lingüísticas*, se muestran las propiedades o características lingüísticas que se han asignado a las respuestas de los informantes⁴.

Es posible combinar varias opciones y atributos a la hora de buscar datos. Así, se pueden consultar:

- Ejemplos recogidos en una sola localidad.
- Ejemplos recogidos en una o más localidades que contengan una o más propiedades lingüísticas.
- Ejemplos proporcionados por informantes de una(s) edad(es) concreta(s) que pertenecen a una propiedad lingüística específica y que han sido recogidos en una o más localidades.
- Etc.

El abanico de posibilidades de consulta que ofrece BASYQUE es, por tanto, muy amplio.

Los resultados de la búsqueda se visualizan en el mapa central (imagen 1). Para ello, se hace uso de las tecnologías que ofrece Google Maps. De este modo, podemos observar cuales son las localidades en las que se detectan las alternancias sintácticas cuyo análisis consideramos de interés. El mapa es una característica propia de los atlas sintácticos y cada usuario puede optar por el tipo de mapa que más le guste: mapa, foto de satélite o la combinación de ambas opciones. Además, las localidades se pueden visualizar dentro de las provincias o las zonas dialectales especificadas por Zuazo (2008).

⁴ A todas las respuestas que recolectamos asignamos ciertas propiedades lingüísticas. Dichas propiedades sirven para indicar el tipo de variación sintáctica que conlleva cada respuesta. Por ejemplo, para el análisis de la variación que existe en lo que se refiere a la concordancia entre el verbo y el dativo, indicamos en cada ejemplo si el argumento dativo está presente o no, el tipo de argumento dativo (nombre propio, nombre común, dativo reflexivo o recíproco...) y si existe o no concordancia con el verbo. O en el caso de la alternancia que ocurre con el evidencial *omen*, especificamos su posición: V + PRT-*omen* + AUX, AUX + PRT-*omen* + V, ... Las propiedades lingüísticas actúan como atributos de búsqueda.

En la parte inferior del mapa, se indica el total de los resultados obtenidos en cada búsqueda y se detalla, en tres ventanas, la información correspondiente a cada una de las respuestas.

En la primera ventana se muestran:

- La pregunta que se ha formulado al informante.
- La traducción de la pregunta al inglés.
- La transcripción de la respuesta que ha proporcionado el informante, tal y como lo ha pronunciado literalmente.
- La escritura formal que corresponde a la respuesta.
- El texto normalizado que corresponde a la respuesta.
- Las glosas (información lingüística detallada) de cada elemento de la frase⁵.
- Las propiedades lingüísticas asignadas a la respuesta.
- Posibles comentarios tanto del informante como del responsable de la encuesta.
- El reproductor de audio para poder escuchar la respuesta proporcionada por el informante.

⁵ Las glosas se asignan automáticamente, gracias al analizador morfosintáctico y un programa informático desarrollados en el grupo IXA de la UPV. El texto normalizado es precisamente necesario para poder glosar las frases automáticamente, ya que de momento no disponemos de analizadores sintácticos que analizan los textos escritos en los dialectos de *Iparralde*. Esta es, sin duda, una aportación muy novedosa a este tipo de Atlas sintácticos, no solo porque facilita el trabajo manual sino también porque toda esta información será una base muy importante para poder crear, en un futuro próximo, herramientas que posibiliten el análisis y procesamiento de las variantes dialectales.

En la segunda ventana se detallan:

- El cuestionario al que pertenece la respuesta obtenida.
- La sección (categoría y subcategoría) dentro del cuestionario al que pertenece la respuesta.
- El tipo de pregunta (traducción, ejercicio de complementación, pregunta de opción múltiple...).
- El responsable de la encuesta.
- Año en el que se ha obtenido la respuesta.
- La meta-categoría y/o la palabra clave (*keyword*) relacionada con esa categoría o subcategoría del cuestionario.

Por último, en la tercera ventana se especifican:

- Los datos del informante (año de nacimiento, sexo y la competencia lingüística en el estándar).
- La localidad a la que pertenece el informante.

Toda la información obtenida por cada respuesta (tanto los ejemplos como los datos correspondientes a los mismos) se puede exportar a documentos de formato pdf, xml o txt. Incluso existe la posibilidad de hacer una exportación avanzada para obtener solamente aquellos datos que más nos interesan (por ejemplo, solamente las respuestas proporcionadas por los informantes eliminando el resto de los datos, o las respuestas con las propiedades lingüísticas asignadas, o únicamente las respuestas con las glosas, o los

cuestionarios conducidos, etc.). Esta posibilidad nos permite poder conservar esos datos en nuestro ordenador y poder utilizarlos para otros estudios.

3.2. Consulta de ejemplos recopilados en otros proyectos

BASYQUE ofrece también la posibilidad de poder consultar ejemplos obtenidos de las conversaciones grabadas en otros proyectos. En realidad, existen varios proyectos en el País Vasco que tienen como objetivo recoger el habla tradicional⁶. En tales proyectos se graban y se recogen conversaciones y/o testimonios de personas que se expresan en su propio dialecto. Normalmente son conversaciones que no están sujetas a un cuestionario específico; por el contrario, en cada grabación los hablantes versan libremente sobre un tema concreto (vacaciones, familia, tradiciones locales, etc.).

En dichas conversaciones se producen ejemplos de variación sintáctica, es decir, los ejemplos afloran sin tener que recurrir a la elicitación de los mismos. Esta nos parece una circunstancia interesante que podemos aprovechar para nuestros estudios. De este modo, además de obtener una mayor cantidad de ejemplos representativos, vamos creando redes de colaboración que nos permitirán ampliar nuestra base de informantes y profundizar en el estudio de las variantes dialectales de la lengua vasca.

En este apartado, las consultas pueden realizarse por medio de los siguientes parámetros diferenciadores: localidades, proyectos, fenómenos lingüísticos⁷, propiedades lingüísticas y por edades de los informantes. Es decir, podemos consultar:

- Ejemplos recogidos en uno o más proyectos.
- Ejemplos que corresponden a una o más localidades.
- Ejemplos que tienen asignada(s) una(s) propiedad(es) lingüística(s) concreta(s).
- Ejemplos proporcionados por hablantes de una o más generaciones.

⁶ *Ahotsak.com* o *EKE.org* son, por ejemplo, dos proyectos importantes que tienen como objetivo recopilar y difundir testimonios orales de vasco-parlantes.

⁷ En el apartado de las conversaciones libres y de los corpus literarios, puesto que los ejemplos recopilados no están ligados a un cuestionario específico, indicamos el fenómeno lingüístico (concordancia, uso del determinante, evidencialidad, nominalización...) que corresponde a cada uno de los ejemplos.

- Etc.

Al igual que ocurre con las respuestas obtenidas mediante los cuestionarios, estos resultados también se visualizan en el mapa, ya que cada conversación esta grabada en una localidad concreta. Y la información correspondiente a cada ejemplo también se ofrece detallada en tres ventanas. En la primera ventana se indica el tema sobre la que versa la conversación, el responsable o el proyecto al que pertenece, el año de la grabación y el vídeo para poder escuchar el testimonio completo.



En esos testimonios grabados podemos escuchar frases características y/o ligadas a la variación sintáctica. Esas frases que nos interesan las transcribimos. En la segunda ventana encontramos una lista de los ejemplos que han sido extraídos y transcritos de la conversación global.



Seleccionando cada ejemplo, uno por uno, podemos ver toda la información que le corresponde: la forma dialectal del ejemplo, su transcripción o escritura formal, el texto normalizado, las glosas, el fenómeno lingüístico que le corresponde, las propiedades lingüísticas asignadas y el audio de la frase transcrita, al igual que en las respuestas obtenidas mediante los cuestionarios.



Finalmente, en la tercera ventana, se detallan los datos que corresponden al informante y a su localidad.



Podemos exportar, además, toda esa información a documentos de formato pdf, xml o txt, mediante los botones que se encuentran en la parte superior de cada ventana.

3.3. Consulta de ejemplos localizados en corpus literarios

Por último, en lo que se refiere a la búsqueda de los ejemplos de variación sintáctica encontrados en corpus o textos literarios, las opciones de consulta disponible son: por autores, por editoriales⁸, por años de publicación, por fenómenos lingüísticos o por propiedades lingüísticas. Así, podemos obtener:

- Solamente los ejemplos localizados en los trabajos de un autor concreto.
- Ejemplos que corresponden a uno o más autores.
- Ejemplos relacionados a fenómenos lingüísticos concretos.

⁸ Existen editoriales (*Maiatz*, entre otras) que tienen como objetivo exclusivo la promoción de la literatura escrita en *Iparralde* sin ninguna exigencia de estandarización.

- Ejemplos que corresponden a uno o más autores y a un año de publicación concreto, y que tienen asignada(s) una(s) propiedad(es) lingüística(s) concreta(s).
- Etc.

La diferencia, en este caso, con las dos anteriores fuentes de información, consiste en que estos ejemplos ni se visualizan en el mapa ni se pueden escuchar puesto que no se trata de ejemplos grabados en una localidad concreta sino que están ligadas a textos escritos. Los resultados y la información correspondiente a cada uno de los ejemplos se detallan en dos ventanas.

En la primera ventana podemos visualizar el ejemplo localizado y su correspondiente información (la forma dialectal del ejemplo, su escritura formal, el texto normalizado, las glosas asignadas automáticamente, posibles comentarios del anotador, el fenómeno lingüístico que le corresponde y las propiedades lingüísticas asignadas).

91 Corpus literario Exportación avanzada PDF XML TXT

Ejemplo localizado en el corpus literario
Zer, jada despeida emaiten dautazu?

Escritura formal Zer, jada despeida emaiten dautazu?

Texto normalizado Zer, jada agurra ematen didazu?

Glosa [Zer]-WHword-ABS [jada]-ADV [agurra]-N-sing-ABS [ematen]-V [didazu]-AUx35(abs)25(erg)15(dat)

Comentario (l, 140)

Fenómenos lingüísticos Dative-agreement

Propiedades lingüísticas Dative - Si
Dative - Si
Agreement - Si
Elided dative - Si

Corpus literario Fuente literaria

Y en la segunda ventana se especifican los datos que corresponden al texto literario en el que se ha localizado el ejemplo (el autor, el título del libro, la editorial, el año de producción, el año de publicación del texto y el área al que pertenece, en caso de haberlo especificado).

1 Fuente literaria Exportación avanzada PDF XML TXT

Fuente literaria Lantzabal, "Lan Guztiak"

Autor Lantzabal, Piarres

Lugar de nacimiento del autor

Título del libro "Lan Guztiak. [Complete Works]"

Año de producción 1934

Editorial Piarres Xarriton. VIII volúmenes. Elkar. Donostia. 1991-1998

Año de publicación

Área geográfica

Corpus literario Fuente literaria

Los datos obtenidos también se pueden exportar a los formatos anteriormente mencionados (pdf, xml y txt).

Por tanto, si bien la fuente principal de información lo constituyen los cuestionarios, BASYQUE también ofrece la posibilidad de importar y consultar ejemplos de alternancia sintáctica que hayan sido localizados en conversaciones grabadas en otros proyectos que trabajan en el campo de la dialectología o en diversos corpus literarios.

De este modo, BASYQUE se convierte en una aplicación muy significativa y valiosa como instrumento de trabajo orientado al análisis de la variación sintáctica que se produce entre los dialectos de una misma lengua.

4. Conclusiones y trabajo futuro

En este artículo hemos presentado BASYQUE, la aplicación desarrollada para el análisis de la variación sintáctica intra-dialectal. Aunque desde nuestra propia perspectiva nos hemos centrado en los dialectos del País Vasco francés, BASYQUE es una aplicación multilingüe que también puede ser útil para el análisis de otras lenguas y dialectos.

Tal como lo venimos señalando a lo largo del artículo, la aplicación nos ofrece la posibilidad de consultar ejemplos que pertenecen a tres fuentes de información: cuestionarios específicos, grabaciones realizadas en otros proyectos y corpus literarios. De este modo, pretendemos i) ofrecer la oportunidad de consultar ejemplos de origen y carácter diferente, ii) aprovechar para nuestros estudios los datos que se han recogido en otros proyectos y textos literarios y iii) crear redes de trabajo en común con los grupos de investigación y fundaciones de tipo cultural que trabajan en el campo de la dialectología.

BASYQUE ofrece, por tanto, diferentes modos de consulta para obtener un tipo de información u otro. El hecho de poseer toda esa información recopilada, organizada y etiquetada en una base de datos, facilita en gran medida los análisis relacionados con la variación sintáctica, ya que de este modo está a disposición de los usuarios una gran cantidad de ejemplos de estructuras sintácticas concretas localizadas geográficamente, clasificadas en relación a determinados parámetros sociolingüísticos y acompañadas de un análisis gramatical básico.

Del conjunto de los datos almacenados, no solo obtenemos ejemplos concretos de alternancias sintácticas sino que, además, tenemos la posibilidad de realizar estudios comparativos, de poder analizar la evolución o la trayectoria por las que han atravesado las diversas variantes dialectales.

Esta aplicación puede resultar, por tanto, un primer paso adelante para aquellos investigadores interesados en el estudio de las variaciones sintácticas en lenguas o dialectos que carecen de bases de datos, plataformas o entornos informáticos que ofrezcan ejemplos e información sobre esas estructuras que sean objeto de estudio. Y en ese contexto encaja perfectamente la situación de los dialectos de *Iparralde*, ya que no existen textos lingüísticamente etiquetados ni bases de datos que faciliten información detallada sobre la variación sintáctica intra-dialectal de esa parte del País Vasco.

Además, toda la información que BASYQUE nos permite almacenar puede servir en un futuro próximo para canalizar el tratamiento automático de las estructuras sintácticas analizadas. Partiendo de la información almacenada en esta aplicación, nuestro primer objetivo es el de abrir nuevas líneas de investigación y dar pasos que permitan progresar en el campo del análisis y procesamiento de las variantes dialectales de *Iparralde*. Para ello, será necesario también estudiar las características léxicas y morfológicas de las variantes dialectales, asociándolos a los ejemplos de las estructuras sintácticas que recogemos en BASYQUE.

Una vez recopilada toda la información arriba referida y tomando como base los sistemas informáticos que el grupo IXA ha desarrollado para el análisis y procesamiento del *euskara*, estaremos en condiciones de crear aplicaciones, tales como marcadores de variantes dialectales, analizadores de textos escritos en diversos dialectos, aplicaciones que posibiliten consultar la forma estándar de una variante dialectal o las posibles variantes dialectales de una forma estándar. Ese tipo de aplicaciones pueden resultar realmente eficaces para seguir investigando no solo en el ámbito de la dialectología sino también en el campo de la lingüística teórica o la enseñanza de idiomas, por mencionar algunas de sus potencialidades.

Desde nuestro punto de vista, BASYQUE es una herramienta práctica que tiene un interés indudable para la comunidad lingüística vasca. Ese interés comprende tres áreas distintas pero

complementarias: por una parte, la constitución de corpus específicos con un soporte informático que den cuenta de la variación lingüística existente en el dominio vasco; por otra parte, la creación de infraestructuras informáticas específicas para el procesamiento de la variación lingüística dialectal; y por último, el desarrollo de bases de datos directamente utilizables para el análisis gramatical de las formas dialectales.

References

- Barbiers and Bennis. 2007. The Syntactic Atlas of the Dutch Dialects. A discussion of choices in the SAND-project. Nordlyd 34, 53-72. Internet: <http://www.ub.uit.no/baser/nordlyd/>.
- Carrilho, Ernestina. 2010. Tools for dialect syntax: the case of CORDIAL-SIN (an annotated corpus of portuguese dialects). In Tools for Linguistic Variation, Gotzon Aurrekoetxea & Jose Luis Ormaetxea (eds.), Anejos de ASJU, LII. Bilbo, UPV/EHU. ISBN: 978-84-9860-429-0.
- Cornips, L. & C. Poletto. 2005. On standardising syntactic elicitation techniques. Part 1. Lingua 115 (7), 939-957.
- Etxepare, Ricardo. 2009. On (some of) the Uses of a Syntactic Atlas. Tinta. E-journal of Hispanic and Lusophone Studies. Department of Hispanic and Portuguese Studies. Santa Barbara, University of California Santa Barbara 8(2), 1-24.
- Etxepare, Ricardo. (aparecerá). The emergence of a dative alternation in north eastern varieties of Basque. En Carme Picallo y Jose María Brucart (eds). *Linguistic Variation in Minimalism*. Oxford: Oxford University Press.
- Etxepare, Ricardo. 2010. *Omen* bariazioan [La partícula evidencial *omen* en sus variantes]. En Beatriz Fernandez, Pablo Albizu y Ricardo Etxepare (eds). *Euskara eta euskarak*. Suplementos de ASJU. Donostia: Universidad del País Vasco-Diputación Foral de Gipuzkoa.
- Fernandez, Beatriz y Ortiz de Urbina, Jon. 2010. Datiboa hiztegian [El dativo en el léxico]. *Euskal Herriko Unibertsitatea*.

Haddican, William. 2005. Aspects of Language Variation and Change in Contemporary Basque. Doctoral Dissertation. New York University.

Holmer, Nils M. 1964. El idioma vasco hablado. Un estudio de dialectología euskérica. Donostia. ASJU.

Ortiz de Urbina, Jon. 1994. Datibo komunztaduraren inguruan [On dative agreement]. In Ricardo Gomez and Joseba Lakarra (eds). Euskal Dialektologiako Kongresua [Congress of Basque Dialectology]. Donostia: Supplements of the International Journal of Basque Language and Linguistics. 579-588.

Zuazo, K. 2008. Euskalkiak, euskararen dialektoak [The Dialects of Basque]. Donostia. Elkar.

Imagen 1: Interfaz de la aplicación BASYQUE.