

Atlas interactivo de la prosodia del Vasco/Euskararen atlas prosodiko interaktiboa, EAPI

Interactive Atlas of Basque Prosody - EAPI

Aitor Iglesias

Universidad del País Vasco, España
Grupo EUDIA, Facultad de Educación de Bilbao,

e-mail: aitor.iglesias@ehu.eus

Aintzane Etxebarria

Universidad del País Vasco, España
Grupo EUDIA, Facultad de Educación de Bilbao,

e-mail: aintzane.etxebarria@ehu.eus

Juan Abasolo

Universidad del País Vasco, España
Grupo EUDIA, Facultad de Educación de Bilbao,
<https://orcid.org/0000-0002-1911-4118>

e-mail: juan.abasolo@ehu.eus

Recibido: 30/11/2020

Aprobado: 02/02/2021

RESUMEN

El *Atlas interactivo de la prosodia del vasco / Euskararen Atlas Prosodiko Interaktiboa* (EAPI) es un proyecto que pretende dar una respuesta multimedia e interactiva a la actual situación de la prosodia en lengua vasca, carente de este tipo de herramientas. El resultado del análisis de este proyecto dará una imagen compleja pero más matizada de lo que es la prosodia en la lengua vasca. Se plantea una profundización teórica, que sobre la base de marcos estructurales que fija la literatura prosódica, se elaboren unos contornos prosódicos adecuados y propios para la variedad estándar, sobre la base de los contornos entonativos más importantes para los diferentes tipos de frases.

Palabras clave: atlas; prosodia; euskera; vasco; interactivo; entonación; acento.

ABSTRACT

The *Interactive Atlas of Basque Prosody / Euskararen Atlas Prosodiko Interaktiboa* (EAPI) is a project that aims to provide a multimedia and interactive response to the current situation of prosody in the Basque language. The result of the analysis of this project will give a complex but more nuanced image of what prosody is in the Basque language. A theoretical deepening is proposed, that on the basis of structural frameworks that prosodic literature sets, suitable and own prosodic contours are elaborated for the standard variety, on the basis of the most important intonative contours for the different types of sentences..

Keywords: atlas; prosody; basque; basque; interactive; intonation; accent

Atlas Lingüístico, según la RAE, es un conjunto de mapas en que se presentan datos lingüísticos procedentes de encuestas.

El primer atlas de la prosodia del idioma vasco constituye una herramienta científica novedosa en el ámbito de la lengua vasca, pues siendo una presentación mapeada de informaciones ya disponibles, al mismo tiempo conserva la masa de datos del corpus como fuente de ulteriores indagaciones, y permite la incorporación constante de nuevas recolecciones, centrados siempre en la distinción precisa de los tres parámetros de la prosodia: 1) frecuencia fundamental (F0), 2) duración, 3) intensidad.

El corpus actualmente está formado por textos, tanto espontáneos como leídos por locutores profesionales y no profesionales, incluye unidades prosódicas con diferente número de sílabas en cada variedad dialectal etiquetados según criterios sociolingüísticos relevantes, además de situados geográficamente.

El *Atlas interactivo de la prosodia del vasco / Euskararen Atlas Prosodiko Interaktiboa* (EAPI: <http://eudia.ehu.es/eapi/>) EAPI está concebido para enriquecerse progresivamente. En ese sentido, están presentes, siguiendo los criterios presentados más arriba, datos provenientes de los diferentes corpus realizados en el grupo de investigación EUDIA el Corpus Oral Dialectal del Euskera / *Euskara Dialektalaren Ahozko Korpusa* (EDAK) (Aurrekoetxea et al., 2009) y el Atlas Sociolingüístico-dialectológico del Euskera / *Euskararen Atlas Soziolinguistiko-dialektologikoa* (EAS) (Aurrekoetxea & Ormaetxea, 2006; Aurrekoetxea, 2010).

La prosodia, como el conjunto de elementos suprasegmentales de la lengua oral, es un campo de discernimiento científico en el que este atlas interactivo se revela necesario.

Hoy día se realizan tres tipos de estudios: el fonológico, el fonético y el de percepción. Por una parte, se describe la posición del acento en las distintas variedades, aun considerando la elevada indeterminación existente en este tema (Aurrekoetxea et al., 2013 & 2014), y se ponderan criterios normativos para la lengua estándar (Gaminde et al., 2014). Por otra parte, se especifican los modelos entonativos más representativos de los principales tipos de oraciones (enunciativos, interrogativos, imperativos, negativos, entre otras) del euskera en las distintas variedades de la lengua. Unido a esto se describen los parámetros prosódicos más relevantes de las actitudes, las emociones y los vocativos expresados en euskera (Gaminde et al., 2017b, Etxebarria et al., 2018). Por último, se acaba de abrir una vía para investigar la prosodia de los diferentes tipos de textos en euskera (Etxebarria et al., 2017, Gaminde et al., 2017a), que momentáneamente resulta precaria debido a que los criterios no están del todo definidos. En un futuro cercano, se espera llevar a cabo una propuesta de definición a ese respecto.

ANTECEDENTES

En el ámbito de la lengua vasca el precursor en representar los datos lingüísticos en un mapa fue Bonaparte (1863); quien, per se, no realizó un atlas, pero sí presentó en un mapa los datos lingüísticos recogidos.

El primer trabajo que se puede considerar como un verdadero atlas lingüístico, a pesar de no llevar la palabra “atlas” en su título, es el *Cuestionario trilingüe* de Azkue (1925) que no se terminó de editar en su totalidad hasta 1984 (Etxaide, 1984). En el *Cuestionario trilingüe* podemos encontrar un total de 350 mapas.

Los primeros trabajos considerados explícitamente como atlas son los de Allières (1978 y 1986) donde se tratan los datos recogidos por Bourciez (1895) y que más tarde también publicarían en su totalidad Aurrekoetxea, Videgain & Iglesias (2004 y 2005). Luego de Allières, cronológicamente le siguieron los dos tomos que realizó la Sociedad de Ciencias Aranzadi (1983 y 1990), en este caso un atlas etnolingüístico.

Lateralmente se pueden mencionar los trabajos de Gaminde (1984, 1985a, 1985b, 1985c, 1985d y 1985e) sobre el verbo del euskera que, si nos ceñimos estrictamente a la

definición de la RAE, también podrían ser tomados en cuenta como atlas lingüísticos. Por último, Gotzon Aurrekoetxea y Xarles Videgain dirigieron el atlas de las hablas populares del euskera realizado por La Academia de la Lengua Vasca (Euskaltzaindia, 2008-2019); el mayor hasta la fecha, que ya cuenta con 10 tomos editados y están en preparación los subsiguientes.

Como se puede observar en este breve resumen de los atlas lingüísticos del euskera, no encontramos ningún atlas prosódico, a pesar de que se publicaron numerosos trabajos sobre la prosodia (Elordieta & Hualde, 2014; Etxebarría et al., 2016; Gaminde et al., 2017a)

El EAPI entonces, cubre un espacio básico e indispensable, hasta ahora vacío en el estudio de la prosodia del euskera. En este contexto, es necesario señalar que las lenguas que cuentan con investigaciones más desarrolladas sobre la prosodia poseen páginas web específicas sobre el acento y la entonación. Podemos referirnos a los ejemplos del portugués¹, de las lenguas románicas², de los dialectos de USA³ o del coreano⁴, entre otros.

En cuanto al acento de palabra, uno de los referentes a nivel internacional es la web *Stress Typ2*⁵, dirigida por Henry van der Hulst y Rob Goedemans; ambos son los máximos impulsores de esta base de datos tipológica que contiene información sobre el sistema métrico de 510 idiomas, entre las que se halla el euskera.

En las lenguas del espacio románico que rodean al euskera, hace ya tiempo que se empezó la construcción de los atlas prosódicos con el proyecto del AMPER (Atlas Multimedia de la Prosodia del Espacio Románico). Como se menciona en su propia página web (Martínez Celdrán & Fernández Planas, 2003-2018), es un proyecto de alcance internacional sobre la prosodia que se fue gestando a finales del siglo XX desde el *Centre de Dialectologie de la Université Stendhal Grenoble* bajo el impulso del Dr. Michel Contini. AMPER contempla el estudio de los tres parámetros prosódicos: frecuencia fundamental (F0), duración e intensidad. Su objetivo principal consiste en estudiar la prosodia de las lenguas románicas y en reflejar los resultados en mapas que se pueden consultar, auditiva y visualmente a través de internet. Este proyecto tiene numerosas ramificaciones⁶ ya que está siendo realizado en varios países de Europa y América Latina. También en esta línea de trabajo, es destacable el *Atles interactiu de l'entonació del català* (Prieto & Cabré, 2007-2012).

El estado actual de los conocimientos sobre la prosodia y en particular sobre la entonación en las lenguas circundantes permite dar pasos seguros en los trabajos sobre la lengua vasca. Así y todo, se ve la necesidad de ahondar en dos aspectos que son el estudio de la prosodia textual y la acción de herramientas interactivas abiertas al público.

OBJETIVOS DEL EAPI

El objetivo general del proyecto consiste en realizar un atlas multimedia del euskera donde se represente gráfica y auditivamente la prosodia de diferentes textos. Al igual que el

¹<http://ww3.fl.ul.pt/LaboratorioFonetica/InAPoP/demo/index.htm>

²<http://prosodia.upf.edu/iari>

³<http://aschmann.net/AmEng>

⁴<http://www.linguistics.ucla.edu/people/jun/ktobi/k-tobi-V2.html>

⁵<http://st2.ullet.net/>

⁶AMPER España e Iberoamérica [<http://stel3.ub.edu/labfon/amper/cast/index.html>]

AMPER-Can (Canarias y Cuba) [<http://labfon.webs.ull.es/proampercan/index.html>];

AMPER-Andalucía y Extremadura [http://www.amprae.es/proyecto_amprae];

AMPER-Cat (catalán y castellano) [http://stel3.ub.edu/labfon/amper/cast/index_ampercat.html];

AMPER-Astur (asturiano, gallego-asturiano y castellano de Asturias) [<http://www.unioviado.es/labo-fone>];

AMPER-Gal (gallego) [<http://ilg.usc.es/amper/>];

AMPER-Port (portugués) [<http://pfonetica.web.ua.pt/AMPER-POR.htm>];

AMPER-Ital (italiano) [<http://www.lfsag.unito.it/amper/index.html>];

AMPER-Rom (rumano) [<http://amprom.uaic.ro/indexamper.php>].

AMPER-Friül (friül) [<http://stel3.ub.edu/labfon/amper/friul/index.html>]

AMPER (Fernández Planas, 2008), el EAPI es un proyecto divulgativo, donde se pone a disposición del público en general una selección de formas representativas de los modelos prosódicos, de diferentes localidades de Euskal Herria, los territorios de habla vasca. Es muy importante que los trabajos sobre prosodia se coloquen en soporte multimedia porque proporciona a los investigadores la prueba empírica necesaria para acceder a la realidad de lo estudiado (Fernández Planas, 2005). El atlas debe proporcionar a los investigadores una base de datos multimedia y un mapa con los puntos geográficos encuestados donde se ofrezca la posibilidad de oír las pronunciaciones de los datos recogidos. El EAPI, entonces, materializado en la creación de la web interactiva, responde a 4 objetivos principales:

- 1) La creación de una página web para albergar el atlas.
- 2) El desarrollo de las técnicas necesarias para la anotación, etiquetado y transcripción de los elementos que conforman el atlas, y la consiguiente selección de materiales que ya disponibles: su anotación, etiquetación y transcripción para incorporarlos al Atlas.
- 3) La recolección de nuevos materiales que enriquezcan el corpus ya a disposición del grupo de investigación EUDIA de la UPV/EHU en las áreas de lengua interesadas, como, por ejemplo, materiales representativos de la producción oral de profesionales de la lengua.
- 4) La búsqueda de patrones lingüísticos geoprosódicos a partir de los parámetros prosódicos, es decir, de duración, intensidad y entonación, que proporcionen a profesionales de la lengua (profesores, locutores, periodistas, actores, etc.) modelos prosódicos (entonativos y acentuales) del vasco, para su aplicación en el desarrollo de sus actividades profesionales.

METODOLOGÍA

En este apartado, vamos a desarrollar cuatro puntos. En primer lugar, se explica cómo se ha creado la página web que alberga el atlas. Posteriormente, se hace un repaso de las diferentes opciones que nos muestra la interface. En tercer lugar, se da a conocer el corpus que conforma este atlas, y en el siguiente punto, se explica cómo se ha realizado la selección de datos. Por último, se expone el modo de hacer la transcripción, etiquetado, segmentación y anotación de datos.

CREACIÓN DE LA WEB

Para la realización de esta página web se conformó un grupo de especialistas multidisciplinar: lingüistas, educadores y educadoras e ingenieros informáticos que diseñaron la estructura de la página (secciones, interactividad, etc.), el tipo de contenidos (acento, entonación, textos espontáneos, textos leídos, etc.), el tipo de objetos a contener (archivos de audio, texto, imágenes) y los formatos de estos (.wav, .txt, .textgrid, .png), más aquellos archivos que futuros usuarios van a poder utilizar en sus propios equipos. Del mismo modo, se realizó un diseño gráfico de la página y su construcción física para albergarla en el servidor del grupo de investigación que se encuentra físicamente en la UPV/EHU. Se consiguió una página web dinámica, que permite al usuario, escuchar los ficheros de audio al tiempo que puede leer los archivos de texto adjuntos y simultáneamente ver los espectrogramas. Por otra parte, también da la opción de descargar los archivos de audio y de texto adjuntos para ser utilizados por los investigadores interesados en este tipo de material.

De modo sucinto, sus características técnicas se basan en el siguiente software de libre de código abierto:

- PHPMailer – Con licencia LGPL
- Bootstrap - Con licencia MIT license
- jQuery - Con licencia MIT license
- Popper.js - Con licencia MIT license
- Leaflet - Con licencia BSD 2-Clause license
- Leaflet-sidebar-v2 - Con licencia MIT licese

- Leaflet.markercluster - Con licencia MIT license
- FontAwesome - Con licencia Font Awesome Free license
- html5shiv - Con licencias MIT y GPL2

LA INTERFACE

La web que acoge al EAPI tiene dos interfaces, una para el usuario público que accede vía internet, y otra para los investigadores autorizados del grupo de investigación EUDIA, que pueden editar e implementar nuevos datos en el atlas.

La web en su página inicial (Figura 1) ofrece una navegación completa a elección entre su parte superior o inferior, desde la que se puede consultar el atlas. En esta pantalla inicial también es posible escoger el idioma de navegación y también, para el caso de los investigadores autorizados, iniciar una sesión de trabajo.



Figura 1. Pantalla inicial del EAPI.

Para un usuario sin permiso adicional, la navegación consta de cuatro páginas. Además de la página inicial tenemos otras tres: la página del atlas, la de su base de datos y la de ayuda. En esta última hay una breve información sobre la utilización de las *cookies* y *JavaScript* y se muestra la información sobre el software de código abierto que se ha utilizado en la realización del sitio web. También se presenta una breve reseña sobre el grupo de investigación EUDIA, y una opción de contacto para eventuales consultas.

La página más visual del EAPI es la referida al propio atlas (Figura 2). En esta página es posible observar un mapa centrado en las zonas vascoparlantes, donde aparecen varios círculos con números que indican de cuáles localidades ya se han incorporado audios, con sus respectivas anotaciones. Los números que aparecen en los círculos indican la cantidad de audios recogidos en el área, localidad o zona señalada. Estos círculos señalan diferentes extensiones de espacio físico según el nivel de zoom que se aplique. El zoom encierra zonas más extensas e información más genérica cuando regulado cerca del cero, y más detallada, sobre una zona menor, cuando se regula sobre números mayores. En esta página, además de poder emplear la herramienta del zoom, también es posible aplicar los filtros que se deseen, en la parte izquierda de la página, para poder visualizar solo los audios que cumplen las características de interés para quien consulta.

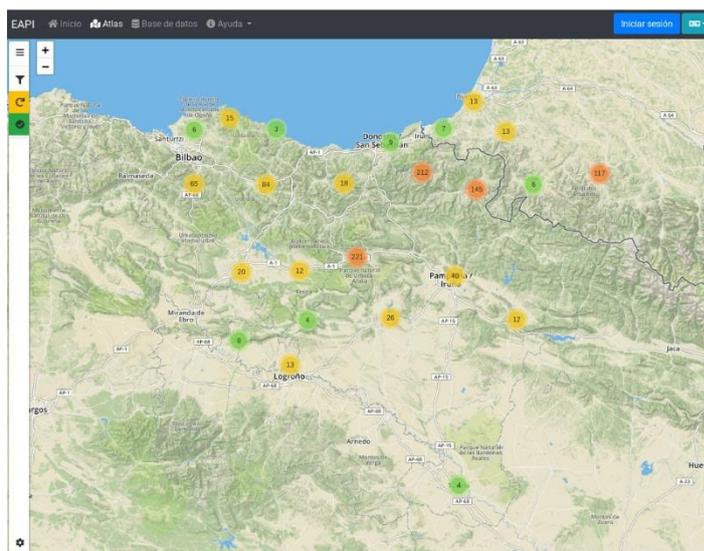


Figura 2. Pantalla del atlas.

Por último, el usuario también puede acceder libremente a la base de datos completa del EAPI (Figura 3). En esta pantalla aparecen todos los datos asociados con cada audio: el informante, el tipo de pregunta, y la localidad. En el caso del informante, se dan los siguientes datos: género, edad, el año de grabación, la lengua materna y si han realizado los estudios en euskera. En el caso de la localidad, se incluye la información de su provincia y también de su región o cantón dependiendo de a qué lado de la frontera corresponda. Las preguntas se diferencian en los siguientes grupos: acento, entonación y texto. En las preguntas sobre el acento se ofrece la información sobre la cantidad de sílabas que tiene y el tipo de sintagma. En el caso de la entonación, se diferencian los tipos de oraciones: enunciativas, preguntas con respuesta afirmativa o negativa, N/Z interrogativa pronominal, optativas, de duda y preguntas eco. Para concluir, se diferencian los textos espontáneos y leídos.

ID	ID del informante	Apellido	Provincia	Región	Cantón	Comarca	Edad	Lengua materna	Estudios en euskera	Año de grabación	Cantidad de sílabas	Tipo de sintagma	Tipo de oración	Oración	Activo	Localidad	Tipo de texto	Género	
1	1	Aceito	Nafarroa	Sabana	Urdabai	Maier	19-25	A - Euskera aprendido en casa	SI	2014	2 Sílabas	SO							
2	1	Aceito	Nafarroa	Sabana	Urdabai	Maier	19-25	A - Euskera aprendido en casa	SI	2014	3 Sílabas	SO							
3	2	Aceito	Nafarroa	Leizola	Gozarta	Maier	19-25	A - Euskera aprendido en casa	SI	2014	3 Sílabas	Talde	Sílabas						
4	2	Aceito	Nafarroa	Leizola	Gozarta	Maier	19-25	A - Euskera aprendido en casa	SI	2014	3 Sílabas	Talde	Sílabas						
5	2	Aceito	Nafarroa	Leizola	Gozarta	Maier	19-25	A - Euskera aprendido en casa	SI	2014	3 Sílabas	Talde	Sílabas						
6	2	Aceito	Nafarroa	Leizola	Gozarta	Maier	19-25	A - Euskera aprendido en casa	SI	2014	4 Sílabas	Talde	Sílabas						
7	1	Tedio	Nafarroa	Sabana	Urdabai	Maier	19-25	A - Euskera aprendido en casa	SI	2014						Común	Leitura	Narración	
8	3	Tedio	Araba	Alabaina	Lauka	Hombre	19-25	B - Euskera aprendido en la escuela	SI	2014						Común	Leitura	Narración	
9	4	Tedio	Araba	Castellu leizola	Castellu	Hombre	19-25	B - Euskera	SI	2014						Común	Leitura	Narración	

Figura 3. Pantalla de la base de datos.

En el caso de los usuarios con permiso de edición, además de las funciones básicas existe la opción de un desplegable (Figura 4) con el que es posible elegir qué tipo de dato editar, ya sea del informante, de la localidad o de la pregunta.

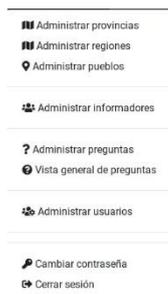


Figura 4. Desplegable del editor.

El EAPI permite añadir cualquier localidad mediante sus coordenadas de latitud y longitud, o bien, obteniendo su geolocalización con un simple clic en el mapa (Figura 5). Es tal la versatilidad del atlas que se podría utilizar para cualquier otro idioma en cualquier otra región del mundo ya que se puede introducir cualquier punto geográfico considerado.

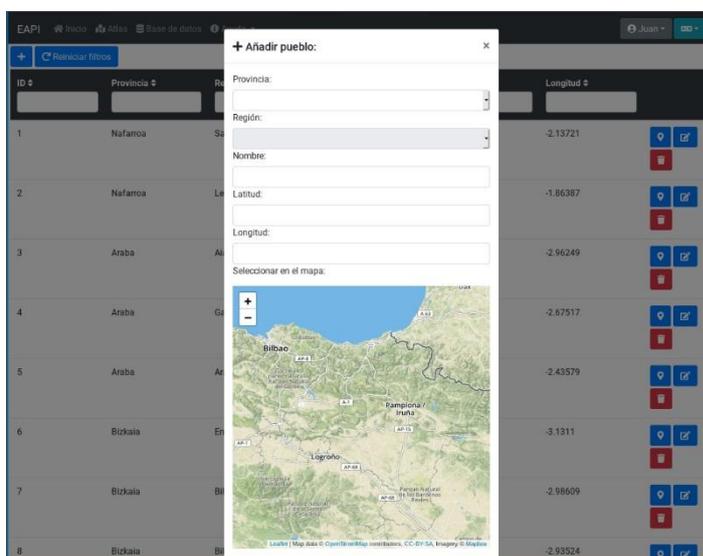


Figura 5. Edición de localidades.

La interface está confeccionada para que cualquier usuario autorizado, con unos mínimos conocimientos informáticos pueda ir alimentando la base de datos. Así, para la inserción de nuevos datos (Figura 6) solo tiene que elegir el informante, el tipo de pregunta, y arrastrar el audio (en formato .wav), y la imagen (en formato .png), creadas con el software Praat.

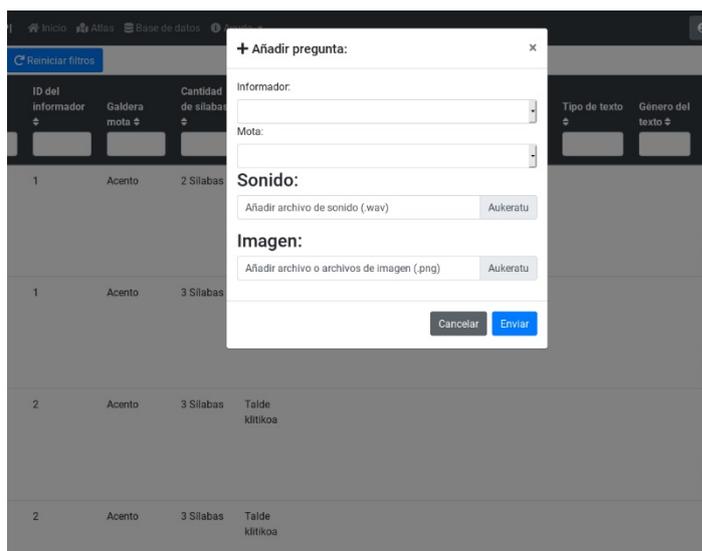


Figura 6. Edición de preguntas.

CORPUS

El grupo de investigación EUDIA de la UPV/EHU, creó dos corpus lingüísticos: el *Corpus Oral Dialectal del Euskera / Euskara Dialektalaren Ahozko Korpusa* (EDAK) (Aurrekoetxea et al., 2009) y el *Atlas Sociolingüístico-dialectológico del Euskera / Euskararen Atlas Soziolinguistikiko-dialektologikoa* (EAS) (Aurrekoetxea & Ormaetxea, 2006; Aurrekoetxea, 2010) que ya han permitido realizar estudios segmentales y suprasegmentales de las variaciones diatópicas, diastráticas y diafásicas.

Aun así, se ha visto la necesidad de seguir recopilando información lingüística y ampliar el corpus teniendo en cuenta los diferentes niveles de la lengua (fonético-fonológico, morfológico, sintáctico y semántico) para poder alcanzar un plan de análisis de la variación lingüística del vasco, incluyendo información sobre distintas generaciones. Para tal fin, la recolección de datos de informantes naturales se lleva a cabo mediante las técnicas usuales en la recogida de datos acerca de la variación lingüística (Aurrekoetxea & Iglesias, 2010).

SELECCIÓN DE DATOS

Una vez construida la estructura de la página web se seleccionaron los datos de entre los que dispone el equipo, procedentes del EAS y del EDAK mencionados anteriormente. Para poder suministrar datos de estos proyectos, en primer lugar, se hizo una selección exhaustiva de los mismos, atendiendo a los criterios elaborados para la construcción de la página, lo que supone análisis lingüístico y análisis estadístico.

Es sabido que la recolección de datos es una fase muy delicada y difícil en toda investigación; sin una buena recogida de datos los resultados de la investigación no son satisfactorios, fiables. Por consiguiente, se procuró, por una parte, que la representatividad de los individuos elegidos fuese alta (teniendo en cuenta el aspecto territorial y el aspecto sociolingüístico); y, por otra, que los datos lingüísticos recogidos representaran la globalidad de los hechos lingüísticos analizados.

La mayor parte de los datos disponibles actualmente por el grupo de investigación corresponden a materiales recogidos para el análisis del acento y de la entonación, tanto con hablantes jóvenes como mayores.

Por otro lado, están también incluidos textos recogidos con hablantes jóvenes, algunos textos espontáneos y algunos materiales sobre la producción de las emociones básicas y las actitudes de los hablantes. Estos datos se sometieron a un examen que permitió seleccionar los datos y las estructuras prosódicas más representativas, desechando aquellas que presentaban una gran

variación de las formas prototípicas. La preparación de esta etapa para conformar el corpus es una de las fases metodológicas más significativas que se llevan a cabo antes de la grabación de los datos (Roseano, 2017). El objetivo es la elección de formas que constituyen el modelo que mejor representa a cada variedad, tanto geográfica como social.

Además de estos datos, de los que ya disponía el grupo de investigación, también se tuvieron que recoger otras muestras significativas del habla, para completar las características más relevantes de la lengua vasca. Estos datos proceden de informantes naturales jóvenes, bilingües de los territorios del País Vasco y se realizaron utilizando diversas técnicas: encuestas con cuestionarios, metodología *task-map*, grabación de habla espontánea y de textos leídos.

Una de los métodos para la recolección de datos es mediante encuestas, con el objetivo de tener respuestas homogéneas de la variación geolingüística. De esta forma, se tiene la certeza de que han sido analizadas todas las variedades del vasco. El hecho de que el área de la lengua vasca dialectal sea de reducido tamaño facilitó de forma considerable la consecución de este fin.

Otra técnica que se utilizó en la recogida de datos es la de *task-map*, o metodología de textos espontáneos, que se llevó a cabo en sus dos vertientes: *map-follower* y *map-giver*.

En la web también hay una muestra significativa de hablas espontáneas, sin ninguna restricción ni condicionante, de unos 5 minutos de duración, y otra muestra de textos leídos. En todas ellas se buscó garantizar la naturalidad de los datos obtenidos, con el objetivo de reflejar un habla real y típica de una comunidad. (Martínez Celdrán & Fernández Planas, 2005). Estos nuevos datos completan los recogidos en proyectos anteriores, de tal forma que el corpus resultante representa un visible enriquecimiento: Además, éstos contribuyen a la conservación de aquellas variedades minoritarias y a conocer cuáles son las de mayor extensión. Estas últimas sirven como base para la posterior postulación de un modelo estándar.

TRANSCRIPCIÓN, ETIQUETADO, SEGMENTACIÓN Y ANOTACIÓN DE DATOS

Una vez hecha la recolección y selección de los datos se procedió a la transcripción, etiquetado, segmentación y anotación prosódica.

A causa de la comparativamente reciente introducción de registros sonoros de la prosodia vasca, existía una gran dispersión de criterios en la transcripción y anotación de los mismos. Este hecho hacía muy difícil la comparación de los datos para poder extraer conclusiones en el plano fonético y fonológico. Además, creaba dificultades en el momento de su posible aplicación en el ámbito de la enseñanza-aprendizaje de la lengua.

En el EAPI no solamente se unificaron los criterios de selección del material recogido en el atlas, sino que, por otra parte, también se ofrecen los mismos criterios de transcripción y anotación en todos los ámbitos del mismo.

Por otra parte, resulta evidente que los modelos prosódicos de lenguas vecinas se están imponiendo a los modelos autóctonos del vasco, incluso cuando los vascoparlantes hablan en su variedad dialectal materna (Etxebarria & Eguskiza, 2018). La realidad de una lengua poco extensa, con modesta incidencia en los medios de comunicación y, por otra parte, la falta de modelos amplios válidos para el euskera estándar incentivan esta situación en la prosodia del vasco.

Como hemos mencionado anteriormente el atlas cuenta con un corpus de prosodia dividido en los siguientes subcorpus:

- Un subcorpus sobre la modalidad del acento en todas las variedades de la lengua incluyendo aquéllas próximas a lo que los hablantes experimentan como variedad estándar.
- Un subcorpus sobre la entonación de los modelos de oraciones básicas (enunciativas con uno y dos sintagmas; interrogativas absolutas, interrogativas parciales, interrogativas con partículas semimodales (*ala*, *ezta*, *ote*).

- Un subcorpus con textos leídos producidos por locutores no profesionales.
- Un subcorpus de textos espontáneos producidos por locutores no profesionales.

El subcorpus sobre el acento cuenta con las entradas (palabras con diferente número de sílabas aisladas y en grupos clíticos, tanto en singular como en plural) suficientes en cada variedad dialectal como para dar cuenta de las reglas fonológicas generales de acentuación de las mismas y poder hacer los correspondientes estudios comparativos entre todas las variedades. También se dan ejemplos de las posibles excepciones de las reglas generales (p.e., palabras marcadas léxicamente y morfemas de plural).

En el subcorpus sobre la entonación de los modelos de oraciones básicas, adecuadamente transcrito y anotado, hay información suficiente para el análisis fonético y fonológico de todas las variedades. En todos los casos, el EAPI cuenta con la información sonora, junto a su transcripción en alfabeto fonético y las correspondientes anotaciones prosódicas. Además, la sección de entonación cuenta con espectrogramas y curvas de registro que ofrecen información de más detalle.

La transcripción de los datos es un momento crítico a la hora de gestionar los diferentes datos (Abasolo et al., 2019). Los datos recogidos mediante grabaciones, ya sean audios o videos, hay que transcribirlos, pero antes se debe decidir qué codificación utilizar. Sabemos que en todas las transcripciones se pierde información (Linell, 2005) porque éstas han de hacerse según el interés del investigador; por lo que es de suma importancia tener muy claro el objetivo de cada trabajo, debido a que el sistema de transcripción tiene que adecuarse a los objetivos de la investigación que se desea realizar (Briz et al., 2002). De todos modos, tenemos que recordar que, aunque se haga la transcripción más concreta y exhaustiva posible, jamás obtendremos toda la información que contiene un archivo de audio. Y de ahí la importancia de que en el atlas prosódico también se pueda recobrar la información sonora original.

En rasgos generales, las transcripciones son de dos tipos: fonológicas o fonéticas. En la dialectología vasca ha sido muy común el uso de las transcripciones fonológicas para textos etnolingüísticos. Las transcripciones fonéticas pueden ser de dos tipos: amplias (también conocidas como aproximadas) o detalladas. Las transcripciones amplias indican solo las características fonéticas más apreciadas de una pronunciación; mientras que las transcripciones detalladas ofrecen más información fonética de los diferentes sonidos. Por lo tanto, cuanto más amplia es una transcripción fonética más se acerca a la transcripción fonológica.

Aunque hay varios alfabetos fonéticos en euskera, el Alfabeto Fonético Internacional (AFI) es el más utilizado. La transcripción de los datos se ha realizado de manera diferente para cada tipo de materiales incluidos en cada sección, de acuerdo con los objetivos de información y características del material de cada sección.

Para los materiales que consisten en palabras o grupos clíticos (subcorpus sobre el acento) se realizó una única transcripción fonológica marcando su correspondiente acento. En el caso de las frases cortas (subcorpus sobre entonación), también se ha hecho una transcripción fonológica; en este caso se han realizado dos tipos más de transcripciones, una según los constituyentes de las frases y otra que contiene la silabificación de los constituyentes. En el caso de los textos, la transcripción fonética tiene cuatro niveles: según los grupos prosódicos del texto, en el nivel de los constituyentes de cada grupo prosódico, en el nivel silábico, y a nivel de los sonidos.

La silabificación se ha realizado de modo semiautomático, por medio de un script para el programa *Praat* (Boersma & Weenink, 2015), programa en el que han sido realizadas todas estas transcripciones, así como los tipos de anotaciones prosódicas que han sido recogidas en las tiras de los ficheros TextGrid que acompañan a cada señal de audio. La segmentación de los datos también se ha hecho de forma semiautomática utilizando y adaptando para ello los scripts del programa; con una posterior corrección manual. Una vez segmentada la señal, se realizó la transcripción en diferentes niveles según el tipo de material y la anotación prosódica.

En esta anotación se han considerado los tipos de tono de frontera, el tipo de pausas distinguiendo las que conllevan alargamiento vocálico o inserción de material sonoro de aquellas que no lo hacen y también se propuso una anotación fonética que considera los picos y los valles de la señal. El sistema de anotación de los datos se adecúa a cada tipo de material (palabras, frases, textos).

CONCLUSIONES Y TRABAJOS FUTUROS

El EAPI es la primera web que recoge la información más relevante sobre la prosodia de la lengua vasca. Es por eso que ha de ser un referente en posteriores investigaciones sobre este tema, ya que pone a disposición de investigadores y público en general una gran base de datos prosódicos para investigaciones con diversas aplicaciones como, por ejemplo, los sistemas de reconocimiento de voz. También ha de ser referente obligatorio para toda la comunidad educativa, puesto que es una herramienta imprescindible para la adquisición del acento y entonación apropiados tanto a nivel del euskera estándar como a nivel dialectal, ya que actualmente hay una considerable degradación en los patrones prosódicos tradicionales del vasco, como se viene demostrando en diferentes investigaciones científicas. Esta degradación tiene su origen, entre otras, en la falta de herramientas para su enseñanza, de modo que esta web se acerca a suplir esta carencia específica.

La comunidad educativa, que ha estado reclamando este tipo de recursos que hagan más fácil y más rápido el aprendizaje de los modelos acentuales y patrones entonativos del vasco, tanto en la educación reglada (desde la primaria hasta la universidad) como en el campo de la enseñanza del euskera como segunda lengua para adultos, tiene ahora una herramienta de la que carecía hasta este momento. El análisis de estos datos permitirá elaborar ejercicios de percepción e imitación de las distintas modalidades del habla. También habrá un impacto social, puesto que además de que esta herramienta pueda ser usada en el aula, también puede ser utilizada en los hogares, con lo que se convierte en soporte complementario de la escuela, con la ventaja de que incluso los referentes en el entorno familiar, educadores y alumnos, tendrán acceso a la web y podrán ayudar a sus hijos en las tareas complementarias.

Este atlas manifiesta también internacionalmente a la lengua vasca en el campo de la prosodia, a la par de otras lenguas, como, por ejemplo, las que se visibilizan dentro del proyecto AMPER. Se une así al conjunto de lenguas más avanzadas en este capítulo de la lingüística. Adicionalmente es aplicable en tecnologías del habla, como síntesis, reconocimiento y sistemas de diálogo. También representa una nueva posibilidad en el tratamiento de patologías del lenguaje, discapacidades, rehabilitación de lesiones que afectan al habla, e incluso para la didáctica de la lengua L1 o L2, en la adquisición de la lengua L1 o L2; y por último en la fonética forense, para la cual esta base de datos es una referencia obligada a la hora de reconocer el origen geográfico de cualquier voz

IMPLEMENTACIÓN DE NUEVOS DATOS

El EAPI no es un trabajo estanco, sino un atlas progresivo, diseñado para incorporar constantemente nuevos datos y nuevos subcorpus. Además de las mencionadas en el punto 3.5. se prevé a breve la recolección nuevas muestras en los siguientes apartados:

- Un nuevo subcorpus con las frases necesarias que recojan los diferentes tipos de emociones básicas en las variedades más representativas (alegría, tristeza, enfado, asco, miedo y asombro).
- Un nuevo subcorpus con las frases necesarias que recojan los tipos de actitudes básicas en las variedades más representativas (ironía, empatía y reproche).
- Agregar al subcorpus de textos leídos grabaciones realizadas por locutores profesionales.

- Agregar al subcorpus de textos espontáneos grabaciones realizadas por locutores profesionales.

ANÁLISIS DE LOS DATOS

Con estos nuevos datos se realizará el análisis lingüístico y estadístico de todo el material referente al subcorpus sobre el acento y al subcorpus de la entonación, paso previo al cartografiado de las áreas más importantes de estos dos aspectos de la prosodia, y se podrá hacer un análisis de las variedades socioprosódicas teniendo en cuenta el género y la lengua materna de los informantes.

Una vez realizado el análisis lingüístico de los datos también se podrá proceder a su análisis estadístico y dialectométrico para a continuación poder igualmente cartografiarlos. El análisis geolingüístico de los datos se llevará a cabo por medio del programa *Diatech*⁷ (Aurrekoetxea et al., 2016), una herramienta web para el análisis y visualización de la variación lingüística que posibilita la gestión de bases de datos, llevar a cabo análisis estadísticos y visualizarlos en gráficos y mapas. Es de suma importancia en la lingüística vasca conocer la implantación geográfica de cada rasgo lingüístico para visualizar cuál es la extensión geográfica de su uso. Para ello, se ha de crear una base de datos con todos los datos lingüísticos extraídos del programa *Praat* y los datos del análisis fonológico. Partiendo de esa base de datos se llevará a cabo el cartografiado automático de los mismos. *Diatech* permitirá al usuario, por una parte, crear un mapa por cada concepto o característica lingüística almacenada en la base de datos, y, por otra, crear mapas sintéticos utilizando diversos procedimientos estadísticos de acuerdo a los diversos intereses clasificatorios. Mediante este cartografiado se visualizarán las áreas prosódico-geográficas más importantes del euskera; cada rasgo lingüístico cartografiado mostrará los datos más significativos y su extensión geográfica. Posteriormente, con el cartografiado de todos los datos del atlas se podrá diseñar una clasificación de las áreas dialectales de acuerdo a la prosodia. La clasificación que se propone se basa en la base de datos prosódicos más extensa disponible; y proporciona a su vez la extensión de dichos rasgos, atendiendo únicamente a hablantes jóvenes, circunstancia que nunca hasta el día de hoy ha sido analizada con profundidad.

Esta gran base de datos que ofrece el EAPI ha de ser utilizada para futuros estudios sobre la prosodia, dialectología, y los resultados han de ser expuestos en dos medios: en la propia web del atlas y otras divulgaciones científicas presentadas en congresos y en publicaciones especializadas.

REFERENCIAS

- Abasolo, J., Iglesias, A., Romero, A. & Unamuno, L. (2020). Distantzia linguistikoa neurtzeko metodoak transkripzio motaren arabera. En J. Abasolo, I. de Pablo & A. Ensunza (eds.), *Hezkuntzari buruzko ekarpenak / Contributions on education* (pp. 96-108). Leioa: UPV/EHU.
- Allières, J. (1978). Petit Atlas linguistique basque français Bourciez I, (20 cartes commentées). *Fontes Linguae Vasconum*, 27, 353-386.
- Allières, J. (1986). Petit atlas linguistique basque français Bourciez II (30 cartes commentées). *Fontes Linguae Vasconum*, 47, 5-45.
- Aranzadi Zientzia Elkarte (1983). *Euskal Herriko Atlas Etnolinguistikoa 1*. Donostia: Aranzadi Zientzia Elkarte.
- Aranzadi Zientzia Elkarte (1990). *Euskal Herriko Atlas Etnolinguistikoa 2*. Donostia: Aranzadi Zientzia Elkarte.
- Aurrekoetxea, G. (2010). Sociolinguistic and Geolinguistic Variation in the Basque language. *Slavia Centralis*, 2010-1, 88-100.
- Aurrekoetxea, G., Gaminde, I., Gandarias, L. & Harignordoqui, E. (2011). EDAK: euskal azentuaren etiketatze bidean lehen emaitzak. En G. Aurrekoetxea & I. Gaminde (eds.), *Prosodiaz eta Hezkuntzaz I. Jardunaldiak/I Jornadas sobre Prosodia y Educación*. Bilbao: UPV/EHU.

⁷<http://eudia.ehu.es/diatech/index>

- Aurrekoetxea, G. & Iglesias, A. (2010). Technology for prosodic variation. En G. Aurrekoetxea & J.L. Ormaetxea (eds.), *Tools for linguistic variation*. (pp. 207-229). Bilbao: Supplements of the Anuario de Filología Vasca 'Julio de Urquijo': LIII.
- Aurrekoetxea, G. & Ormaetxea, J. L. (2006). Research project "Socio-geolinguistic atlas of the Basque language". *Euskalingua*, 9, 157-163.
- Aurrekoetxea, G., Sánchez, J. & Odrizola, I. (2009). EDAK: A Corpus to Analyse Linguistic Variation. En P. Cantos & A. Sánchez (eds.), *A Survey on Corpus-based Research. Panorama de investigaciones basadas en corpus*. (pp. 489-503). Murcia: Asociación Española de Lingüística del Corpus.
- Aurrekoetxea, G., Santander, G., Usobiaga, I. & Iglesias, A. (2016). Diatch: Tool for making dialectometry easier. *Dialectología*, 17, 1-22.
- Aurrekoetxea, G., Videgain, X., & Iglesias, A. (2004). 'Bourciez' bildumako Euskal Atlas: 1. Lexikoa. ASJU XXXVIII-2.
- Aurrekoetxea, G., Videgain, X., Iglesias, A. (2005). 'Bourciez' bildumako Euskal Atlas: 2. Gramatika. ASJU XXXIX-1.
- Aurrekoetxea, G., Gaminde, I., Iglesias, A. & Gandarias, L. (2013). Prosodic Variation in the Basque Language: Stress Areas. En E. Carrilho, C. Magro & X. Álvarez, (eds.), *Current Approaches to Limits and Areas in Dialectology*. (pp. 247-266). Newcastle: Cambridge Scholars Publishing.
- Aurrekoetxea, G., Gaminde, I., Gandarias, L. & Iglesias, A. (2014). Prosodic variation in the Basque language: intonational areas. En A. Díaz-Negrillo & F.J. Díaz Pérez (eds.), *Specialisation and Variation in Language Corpora*. Berna: Peter Lang AG, International Academic Publishers.
- Boersma, P. & Weenink, D. (2020). *Praat: doing phonetics by computer*. (Computer program., version 6109). Consultado el 27 de febrero del 2020 en <http://www.praat.org/>.
- Bonaparte, L.L. (1863). *Carte des sept Provinces Basques*. London: Stanford's Geographical Establishment.
- Bourciez, E. (1895). *Recueil des idiomes de la région gasconne*. Eskuizkribua, Bordeaux: Bibliothèque universitaire des lettres.
- Briz, A. & Grupo Val.Es.Co (2002). *Corpus de conversaciones coloquiales*. Madrid: Arco/ Libros.
- Elordieta, G. & Hualde, J.I. (2014). Intonation in Basque. *The phonology of intonation and phrasing*, 405-463.
- Etxaide, A.M. (1984). *Erizkizundi hirukoitza / Triple cuestionario / Triple questionnaire (Euskera 1925)*. IKER 3. Bilbo: Euskaltzaindia.
- Euskaltzaindia, (2008-2019). *Euskararen Herri Hizkeren Atlas* (10 tomos). Bilbo: Euskaltzaindia.
- Etxebarria, A. & Eguskiza, N. (Eds.) (2018). *Bariazioa esaldien intonazioan*. Bilbo: UPV/EHU.
- Etxebarria, A., Eguskiza, N., Gaminde, I., Romero, A. (2018). Ahots seduzitzaileen ezaugarriak generoaren ikuspegitik. En L. Unamuno, A. Romero, A. Etxebarria & A. Iglesias (eds.), *Linguistic variation in the Basque language and Education/Euskararen bariazioa eta bariazioaren irakaskuntza* (pp. 43-59). Leioa: UPV/EHU.
- Etxebarria, A., Gaminde, I., Romero, A. & Iglesias, A. (2016). Desarrollo de la competencia prosódica en la lectura en voz alta: importancia de las pausas. *OCNOS Revista de estudios sobre lectura*, 15, 110-118.
- Etxebarria, A., Sanz, A., Garay, U. & Romero, A. (2017). Unibertsitateko ikasle elebidunen prosodia gaitasuna Txanogorritxu ipuinaren kontaketa: muga-tonuak. En A. Iglesias & A. Ensunza (eds.), *Gotzon Aurrekoetxea lagunarterik hara* (pp. 111-122) Bilbo: UPV/EHU.
- Fernández Planas, A.M. (2005). Datos generales del proyecto AMPER en España. *Estudios de Fonética Experimental*, 14, 13-27.
- Fernández Planas, A.M. (2008). El proyecto Atlas Multimèdia de la Prosòdia de l'Espai Romànic (AMPER) i les "III Jornades científicas del proyecto AMPER" (24-25 d'octubre de 2006). *Estudis Romànics*, Vol. 30, 233-239.
- Gaminde, I. (1984). *Aditza Bizkaieraz I*. Pamplona: UEU.
- Gaminde, I. (1985a). *Aditza Bizkaieraz II*. Pamplona: UEU.
- Gaminde, I. (1985b). *Aditza Bizkaieraz III*. Pamplona: UEU.
- Gaminde, I. (1985c). *Aditza Ipar Goi Nafarreraz I*. Pamplona: UEU.
- Gaminde, I. (1985d). *Aditza Ipar Goi Nafarreraz II*. Pamplona: UEU.
- Gaminde, I. (1985e). *Aditza Ipar Goi Nafarreraz III*. Pamplona: UEU.
- Gaminde, I., Aurrekoetxea, G., Etxebarria, A., Garay, U. & Romero, A. (2014). *Ahoskera lantzeko argibideak eta jarduerak. Laguntzarako materiala: teoria eta praktika*. Leioa: UPV/EHU.
- Gaminde, I., Etxebarria, A., Romero, A. & Eguskiza, N. (2017a). Características de la competencia prosódica de jóvenes bilingües vascos en la lectura en voz alta: las cumbres tonales. *Revista de Lingüística Teórica y Aplicada*, 55, 35-72.
- Gaminde, I., Iglesias, A., Eguskiza, N. & Etxebarria, A., (2017b). Tasun akustikoen eragina jarrerren bereizkuntzan. *Ikastorratza*, 19, 126-149.
- Linell, P. (2005). *The written language bias in linguistics. Its nature, origins and transformations*. Londres: Routledge.

- Martínez Celdrán, E. & Fernández Planas, A.M. (coords). (2003-2018). *Atlas Multimèdia de la Prosòdia de l'Espai Romànic*. Consultado el 24 de febrero del 2020 en http://stel.ub.edu/labfon/amper/cast/index_ampercat.html
- Martínez Celdrán, E.& Fernández Planas, A.M. (2005). Estudio metodológico acerca de la obtención del corpus fijo en el proyecto AMPER. *Estudios de fonética experimental*, 14. 29-66.
- Prieto, P. & Cabré, T. (coords.) (2007-2012). *Atles interactiu de l'entonació del català*. Consultado el 10 de febrero del 2020 en <http://prosodia.upf.edu/atlesentonacio>.
- Roseano, P. (2017). Notes metodològiques sobre la construcció del corpus fix d'AMPER en el cas d'una llengua sense estàndard oral. *Estudios de Fonética Experimental*, 26, 283-309.