

ANÁLISIS PRELIMINAR EN DIFERENTES NIVELES SINTÁCTICOS DE LA RETROCARGA EN EUSKERA^{1*}

Mikel Iruskietia

Euskal Herriko Unibertsitatea (UPV/EHU). IXA Group

Julian Maia-Larretxea

Euskal Herriko Unibertsitatea (UPV/EHU). DREAM Research Group

RESUMEN

La discusión sobre cuáles son las reglas sintácticas del euskera relativas al orden de colocación de los elementos de la oración es un debate con largo recorrido y ya hay diferentes propuestas relacionadas con este tema, en particular, sobre la retrocarga (aceptadas en mayor o menor grado y en uso); aunque no conocemos estudios que describan el grado de retrocarga en función de diferentes niveles sintácticos y la influencia de algunos factores sociolingüísticos de los usuarios y de las diferentes variables que pueden afectar. Presentamos en este artículo un estudio preliminar de la retrocarga en correcciones sintácticas del alumnado de bachiller en euskera y observamos diferentes variables que pueden influir en cómo se concibe la estructura sintáctica en euskera. Para realizar este estudio hemos encuestado a 114 estudiantes de 5 centros de bachillerato de la EAE/CAV de modelos lingüísticos diferentes, que han corregido 45 enunciados largos con una retrocarga mínima. Los resultados obtenidos son prometedores y señalan en qué nivel sintáctico se produce una mayor retrocarga y el comportamiento de algunas variables sociolingüísticas que en ello influyen.

Palabras clave: orden de palabras, subordinación, sintaxis, retrocarga, euskera

^{1*} Agradecemos a Jon Alkorta y Unai Atutxa del Grupo IXA por la labor de anotación de la retrocarga. La anotación y su evaluación han sido presentados en el trabajo de Iruskietia et al. (2016). Este trabajo ha sido financiado por el proyecto “Investigando el orden de las palabras en la prosa lógico-discursiva escrita en euskera” (UPV/EHU).

ABSTRACT

There is a long discussion about the syntactic rules in Basque language related to word order in the sentence and there are different proposals related with this topic, in particular, about the rear-burden (accepted more or less and in use by different users); and, even though, as far as we know there is not any study which describe the different grades of the rear-burden in different syntactic levels and the influence of different sociolinguistic variables. This paper presents a preliminary study about the rear-burden of syntax corrections of bachelor students in Basque written texts and we conclude that different variables can affect in how students consider the syntactic structure in Basque language. To do so, we administered a questionnaire to 114 students in 5 different high-schools with different language models in the Autonomous Community of the Basque Country. The respondents had to correct 45 sentences with a minimum rear-burden. The obtained results are promising and they show in which syntactic levels the rear-burden is greater and how some sociolinguistics variables affect.

Keywords: word order, subordination, syntax, rear-burden, Basque language

1. INTRODUCCIÓN

El estudio del orden sintáctico de la oración en euskera y la retrocarga proveniente de ciertas reglas sintáctico-gramaticales o las correspondientes estrategias para la disminución de la retrocarga es un tema de preocupación que viene de lejos (Altube 1929) y ha suscitado gran interés por parte de diferentes autores. Maia (2014) realiza un resumen y una clasificación de una serie de autores significativos que han tratado el tema desde diversos puntos de vista. Véanse, entre otros, los trabajos de Altube (1975), Hidalgo (2002), García (1996), Garzia (2014), Alberdi y Sarasola (2001), Rubio (2002), Kaltzakorta (2012), Aristegieta (2012), Esnal (2011), Amuriza (2012), Agirre (2013) y Sarasola (2016) por citar algunos.

Para estimar la aceptación de la retrocarga entre los estudiantes de Bachillerato de la *Euskal Autonomia Erkidegoa*/Comunidad Autónoma Vasca (EAE/CAV), Maia (2015) e Iruskietia *et al.* (2016) han computado la forma de corregir la retrocarga por parte de expertas/os y por muestras del alumnado tanto de grado universitario como de bachillerato y los datos obtenidos en esos estudios señalan que hay ciertas diferencias en cuanto a la corrección de la retrocarga.

Por ejemplo, se presentó a los y las encuestadas la oración (1a) y se solicitó su corrección en caso de que consideraran no aceptable el orden de sus elementos.

(1) a. **Beldur naiz** orain arte egin ditugun lan guztiak alferrekoak izango direla.

Me temo que todos los trabajos que hemos hecho hasta ahora sean en vano.

b. Orain arte egin ditugun lan guztiak alferrekoak izango diren/izango direlako/*direnaren **beldur naiz**.

Todas y todos los integrantes del grupo de expertos dieron por bueno el orden de los elementos propuesto inicialmente en (1a) con el verbo principal *beldur naiz* ‘me temo’ al principio de la oración, mientras que un 66% de las y los estudiantes de grado universitario lo corrigieron, optando por varias alternativas que tenían en común la característica de tener el verbo principal al final de la oración, como aparece en (1b).

El objetivo de nuestro trabajo es analizar si las correcciones de la retrocarga sobre oraciones largas en euskera por parte de alumnos de bachillerato difieren dependiendo de la estructura sintáctica a corregir y de alguna otra variable sociolingüística que incida significativamente. Este análisis nos permitirá ob-

servar si hay diferencias a la hora de corregir la retrocarga entre diferentes niveles sintácticos o entre el alumnado de diferentes características. No tenemos conocimiento de la existencia de trabajos previos que traten de la corrección de diferentes niveles sintácticos o de las características que pueden influir en la retrocarga.

En la sección 2 damos cuenta de la metodología empleada en nuestro estudio y en la sección 3 mostramos los resultados obtenidos en nuestro análisis. Finalmente, en la sección 4 exponemos las conclusiones, limitaciones y trabajo futuro.

2. METODOLOGIA

2.1 Obtención de datos

El corpus de estudio está formado por correcciones sobre la retrocarga en 114 encuestas realizadas por estudiantes de bachiller en 5 centros diferentes de la EAE/CAV, en las que cada estudiante ha respondido a 45 enunciados, extraídos de textos reales, con un promedio de 14 palabras por oración (la oración más corta tiene 9 palabras y la más larga 21).

Cada estudiante ha tenido aproximadamente una hora para leer y realizar las correcciones oportunas en papel. Todos los y las estudiantes han tenido tiempo para terminar las correcciones. Los enunciados han sido diseñados con la menor retrocarga posible, aunque hemos hecho algunas excepciones con estructuras inaceptables para saber si los y las estudiantes están respondiendo con la atención debida. A los y las estudiantes se les ha pedido que corrijan las oraciones en el caso de que las consideren no-gramaticales.

A continuación presentamos en el ejemplo (2) ese tipo de oraciones que se espera que sean corregidas por la totalidad o la inmensa mayoría de las personas encuestas. La oración, colocada hacia la parte final del conjunto de oraciones presentadas para su análisis, contiene una forma compacta o no-perifrástica del verbo que ha sido situada en la posición inicial absoluta de la oración:

(2) a. ***Nago** oso nekatua, ez naiz irtengo.

Estoy muy cansado, no voy a salir.

Esa ordenación de los elementos fue corregida por más del 95% de las personas encuestadas, lo que indica un muy buen nivel de atención correctora por su parte.

Aunque las encuestas han sido anónimas, se ha solicitado a los y las estudiantes algunos datos de orden sociolingüístico que permitan plantear hipótesis sobre los resultados observados, como por ejemplo: género, entorno lingüístico de procedencia, modelo lingüístico, edad, titulación en lenguas extranjeras, y proporción estimada de uso de las lenguas por parte del estudiante.

Resumiendo, las características, a nuestro entender, más importantes son las siguientes: han sido analizados datos provenientes de un total de 5 institutos de Enseñanza Secundaria Post-obligatoria diferentes y 114 estudiantes (6% de Araba, 70% de Bizkaia y 22% de Gipuzkoa con una media de edad de 16,8 años), provenientes y usuarios de modelos lingüísticos diferentes (75% de estudiantes en modelos de inmersión en euskera, y 23% en modelos bilingües parciales) en el año 2015. De todos ellos el 63% han señalado que son mujeres, 31% hombres y un 6% ha marcado otro género.

Podemos ver en la Figura 1 la información lingüística de los y las estudiantes: la primera lengua es para el 17% el euskera, para 4% el euskera y el castellano, para el 77% el castellano y un 3% tiene otra lengua materna. Los estudiantes utilizan las lenguas en la siguiente proporción: el 30% del tiempo el euskera (mayoritariamente en tiempo escolar), el 64% el castellano y el 6% otra lengua.

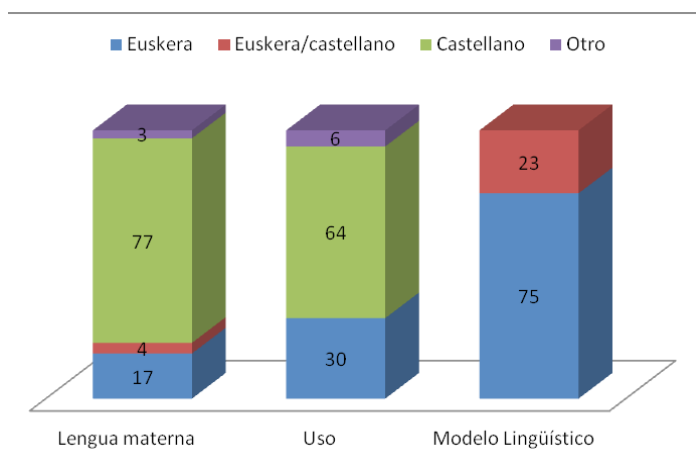


Figura 1. Descripción de los estudiantes que han participado en el estudio

2.2 Fenómenos lingüísticos a observar

La anotación del peso de las correcciones se ha llevado a cabo siguiendo una guía elaborada por los autores, donde las correcciones se ponderan de una forma determinada. Cuando se acepta la ordenación presentada el ítem se valora como 0 (cero = ausencia de retrocarga) y si se corrige y se coloca al final de la oración se puede valorar hasta 1 (retrocarga máxima); en medio hay diferentes valores que dependen de la posición intermedia a la que se ha llevado el ítem en cuestión (véase ejemplos (3), (4), (5) y (6)).

Para conocer el grado de fiabilidad de esta guía el proceso de anotación se ha desarrollado por parte de tres personas, que trabajaban de manera aislada, sin consultas mutuas. Los tres anotadores han trabajado sobre las mismas encuestas (9 encuestas, equivalentes a un 8% del corpus total analizado; calculamos que este trabajo ha costado a cada anotador entre 6 y 8 horas de dedicación). El grado de fiabilidad por pares en términos de F1 (*f-measure*) ha sido de entre un 0.926 (siguiendo la guía estrictamente) y un 0.862 (cuando uno de los anotadores ha seguido un modo más intuitivo) (Iruskieta *et al.* 2016).

Una vez que se ha evaluado la fiabilidad de la guía de anotación y de los anotadores, se han anotado las 105 encuestas restantes (anotar cada una de las encuestas supone una actividad de entre 20 y 30 minutos), para poder responder a las dos preguntas principales de este artículo, que son las siguientes:

i) ¿Las correcciones sobre la retrocarga se realizan de igual modo en los diferentes niveles sintácticos?

ii) ¿Qué variable sociolingüística altera significativamente la retrocarga?

Para responder a esas preguntas hemos diferenciado los 61 ítems de estudio en cuatro niveles sintácticos siguiendo la degradación jerárquica de la combinación entre oraciones (*hierarchical downgrading*, Lehman 1985): i) oración principal, ii) oración subordinada adjunta con verbos personales, iii) oración subordinada adjunta con verbos impersonales y iv) oraciones complementarias.

Por lo tanto, en algunos casos tenemos más de un ítem de observación de diferentes niveles sintácticos en un solo enunciado, ya que en los 45 enunciados a corregir vamos a observar el comportamiento de 61 ítems.

En los siguientes apartados comentamos brevemente con ejemplos los cuatro niveles sintácticos que hemos analizado en este estudio.

2.2.1 Oración principal

Como mayoritariamente los enunciados tienen inicialmente una retrocarga mínima, en este nivel de análisis queremos observar si los estudiantes corrigen el verbo principal retrasándolo hacia el final de la oración, provocando de esa manera un aumento de la retrocarga.

Como ejemplo presentamos en qué medida y por qué se corrige el ejemplo (3) y cómo se aumenta la retrocarga.

(3) Atzo eztabaida **izan nuen** lagun batzuekin ___ munduan gertatzen ari diren gerra batzuei buruz ___.

*Ayer discusión **tuve** con algunos amigos ___ sobre varias guerras que están sucediendo en el mundo ___.

Concretamente queremos ver si el estudiante corrige el verbo principal con su auxiliar *izan nuen* ‘(yo) tuve’ hacia uno de los huecos que hemos puesto en el ejemplo, teniendo en cuenta que esa corrección dificulta la lectura y la comprensión del enunciado. Las dos posiciones en que se puede colocar el verbo en el caso de procederse a la corrección de su formulación inicial se han marcado de la siguiente forma (mediante una barra baja continua): ‘___’, representando así las posiciones en las que esperamos que los y las estudiantes pongan el verbo.

Si bien es cierto que los estudiantes encuestados pueden poner el verbo en otra posición manteniendo la gramaticalidad, por claridad explicamos aquí sólo los dos lugares más comunes hacia los que se espera que se produzca el desplazamiento. De estas tres posiciones, la original se ha anotado con peso 0 y las demás una con un 0,5 en una posición intermedia y 1 en posición final.

2.2.2 Oración subordinada adjunta con verbos personales

El siguiente nivel que queremos observar son las oraciones subordinadas adjuntas con verbos personales como en el ejemplo (4). En esta ocasión esperamos que el sintagma verbal *badakien arren* ‘aunque (él/ella) sabe que’ sea retrasado hacia el final de la oración.

(4) Ez du batere ahaleginik egin guri laguntza bidaltzeko, **badakien arren** oso jende gutxi garela eta garaiz amaitzeko lan handiegia dugula ___.

Él/Ella no ha ningún esfuerzo hecho para mandarnos ayuda, **aunque él/ella sabe que** somos poca gente y que tenemos demasiado trabajo para terminar a tiempo ____.

2.2.3 Oración subordinada adjunta con verbos impersonales

También queremos analizar qué sucede con las oraciones subordinadas adjuntas que tienen verbos impersonales. Así, queremos estudiar si en el ejemplo (5) también *ikusteko* ‘para ver’ será retrasado hacia al final de la oración.

(5) Hiru irakasleak gaiari ikuspegi historikoa ematen saiatu dira, **ikusteko** nola garatu diren euskarazko hedabideak azken 10 urteotan ____.

Los tres profesores han intentado dar una perspectiva histórica al tema, **para ver** como se han desarrollado los medios de comunicación en euskera en los últimos 10 años ____.

2.2.4 Oraciones complementarias

Por último, queremos ver si sucede lo mismo en las oraciones complementarias como en el ejemplo (6), corrigiendo la palabra *berria* ‘noticia’ y poniéndolo al final de la oración. En este caso, la corrección necesita de un cambio morfosintáctico. Pero si el alumnado encuestado piensa que el núcleo de la oración complementaria deba colocarse atrás, realizará ese cambio, con lo que aumentará la retrocarga.

(6) Telebistan entzun dut **berria** aurten Tourrean kontrol zorrotzagoak jarriko dituztela ____.

He oído en la televisión **la noticia** de que van a poner controles más selectivos en el Tour de Francia.

3. RESULTADOS

Para observar qué estructura sintáctica se ha corregido en mayor medida, ofrecemos en la Tabla 1 los datos generales de la retrocarga en los diferentes niveles de análisis. En primer lugar, presentamos los datos generales de la retrocarga; en segundo lugar, los datos sobre los fenómenos de corrección que afectan a la

oración principal y a las oraciones subordinadas; y, por último, presentamos los datos de las oraciones subordinadas en tres niveles diferentes con su número de ítems y el peso total de la retrocarga: i) oraciones subordinadas personales, ii) oraciones subordinadas impersonales y iii) oraciones complementarias. Se han calculado dos medidas de la retrocarga: i) retrocarga media por persona (retrocarga total/personas encuestadas) y ii) retrocarga media por ítem (retrocarga total/ítems a corregir).

	General	O. princ.	O. subord	O. sub. person.	O.sub. imperson.	O. compl.
Ítems a corregir	61	28	32	6	16	11
Retrocarga total	2942,50	1260,30	1717,00	297,50	935,25	449,50
Retrocarga media por persona	25,59	12,75	14,93	2,59	8,13	3,91
Retrocarga media por ítem	48,23	45,01	53,66	49,58	58,45	40,86

Tabla 1. Datos generales de la retrocarga en cada nivel de estudio

Los resultados obtenidos indican que el grado de corrección es diferente en los diferentes niveles sintácticos. En primer lugar, queremos observar si la diferencia de corrección entre las oraciones subordinadas (1717) que se corrigen algo más que las oraciones principales (1260,3) es significativa. Para ver esta diferencia entre estos dos grupos, hemos usado la prueba t de Student (t-test) para muestras independientes asumiendo un p-valor menor que 0,05 para refutar la hipótesis nula (que no hay diferencia significativa cuando se corrigen diferentes niveles sintácticos), ya que los datos tienen una distribución normal según el test de Kolmogorov-Smirnov (K-S) (el p-valor del test K-S ha sido en este caso menor que 0,001). Puesto que el p-valor del t-test es de 0,001 (asumiendo o no varianzas iguales), es decir, menor que 0,05, rechazamos la hipótesis nula y concluimos que sí hay diferencias de corrección en función de los diferentes niveles sintácticos.

Para observar si el peso promedio difiere o hay diferencias entre las correcciones realizadas en diferentes niveles sintácticos hemos realizado un análisis *post-hoc* con el test de Tukey en la Tabla 2.

(I)	(J)	Diferencia de medias (I-J)	Error típico	Signif. p-valor	Intervalo de confianza al 95%	
					Límite inferior	Límite superior
O princ.	O. sub. pers.	844,52	89,82	0,000	612,90	1076,13
	O. sub impers	178,95	89,82	0,192	-52,66	410,56
	Compl.	919,84	89,82	0,000	688,22	1151,45
O. sub. pers.	O. sub impers	-665,57	89,82	0,000	-897,18	-433,96
	Compl.	75,32	89,82	0,836	-156,28	306,94
O. sub impers.	Compl.	740,89	89,82	0,000	509,28	972,50

Tabla 2. Test de Tukey para comparar diferencias de la retrocarga entre niveles sintácticos

Gracias al test de Tukey de la Tabla 2 podemos ver que no todas las combinaciones son significativas, ya que el p-valor es superior a 0,05, en el caso de la comparación entre las oraciones principales y las oraciones subordinadas impersonales (el p-valor es de 0,192) o entre las oraciones subordinadas personales y las oraciones complementarias (el p-valor es de 0,836).

Resumiendo todos esos datos observamos en la Tabla 3 que sí hay diferencias significativas cuando se comparan los niveles sintácticos de la tercera con los de la cuarta columna (por ejemplo Compl. con O. princ), pero no hay diferencias significativas (p-valor superior al 0,05) entre los niveles que se encuentran en la misma columna (por un lado, compl. con O. sub. pers. y, por otro lado, O. sub. impers. con O. princ.).

Estructuras sintácticas	N	Subconjunto para alfa = 0.05	
		1	2
Compl.	114	185,64	
O. sub pers.	114	260,96	
O. sub impers.	114		926,54
O. princ.	114		1105,48
Significancia		0,836	0,192

Tabla 3. HSD de Tukey para comparar diferencias en el promedio de la retrocarga entre diferentes niveles sintácticos con un tamaño de muestra de 114 estudiantes.

También queríamos analizar si había correlación entre la degradación jerárquica (*hierarchical downgrading*) y la corrección de la retrocarga. En ese caso, esperábamos encontrar mayor retrocarga en las oraciones principales y que la corrección fuese menor a medida que se va degradando la jerarquía sintáctica, con la correspondiente disminución de la retrocarga. Pero los datos no apuntan en esa dirección, ya que las oraciones subordinadas se corrigen más que las oraciones principales.

Otra de las cuestiones interesantes era observar el comportamiento de otras variables sociolingüísticas que podrían afectar en la retrocarga. Este análisis también se ha realizado con el t-test y presentamos los resultados en la Tabla 4. Las variables que hemos estudiado son las siguientes: el modelo lingüístico (hemos comparado el modelo de inmersión o modelo D con el modelo en castellano y el bilingüe parcial, modelos A y B), el nivel de competencia en otras lenguas (acreditada por titulación oficial), la variable del género o de la edad.

	Modelo lingüístico		Titulación en otras lenguas		Género		Edad	
	Inmers.	Otros	Sí	No	F	M	17/18	15/16
Valor medio	20,91	4.89	15,23	6,00	17,37	7,61	5,24	20,57
Difer.	16,02		9,23		9,76		-15,33	
p-valor	0,003		0,000		0,175		0,007	

Tabla 4. Variables sociolingüísticas de estudio sobre la corrección de la retrocarga media.

Según la Tabla 4 cuanto más se usa o conoce el euskera o se tenga competencias también en otras lenguas, la retrocarga es mayor en la corrección. Las variables que más influyen y que son significativas son las del modelo lingüístico (diferencia de 16,02 con un p-valor inferior a 0,05) y la edad (diferencia de 15,33 con un p-valor inferior a 0,05) y, en un grado menor, conocimiento de otras lenguas (diferencia de 9,23 con un p-valor inferior a 0,05). El género, aunque no es la variable con menor diferencia (9,76) resulta una diferencia no significativa (ya que tiene un p-valor superior al 0,05), mientras que la variable de la titulación en otras lenguas (9,23), sí lo es (p-valor inferior a 0,05).

4. CONCLUSIONES Y TRABAJO FUTURO

En este estudio hemos analizado el grado de la retrocarga en las correcciones hechas por estudiantes de bachiller entre 15 y 17 años en la EAE/CAV y hemos observado que la retrocarga es mayor en las oraciones principales que en las subordinadas y que hay diferencias a la hora de corregir los diferentes niveles sintácticos. Además, hemos observado cuáles pueden ser algunas de las razones sociolingüísticas que pueden afectar a la retrocarga, siendo el modelo lingüístico y la edad las variables con mayor incidencia.

En cambio, cuando la corrección ha sido hecha por expertos o expertas, los resultados apuntan en la dirección contraria, ya que en este caso, un mayor grado de conocimiento de la lengua apunta hacia una retrocarga menor, tal y como apuntan Maia y Larrea (2013). Por ello, se necesita profundizar en la investigación de la retrocarga para conocer realmente las razones de estos comportamientos aparentemente divergentes.

Así pues, en el futuro tenemos intención de analizar las respuestas de más estudiantes de bachiller y otros fenómenos lingüísticos o variables socioculturales que podrían afectar a la retrocarga en euskera.

REFERENCIAS BIBLIOGRÁFICAS

- ALTUBE, S. 1929. *Erderismos*. Gaubeka Irarkola. Bermeo.
- AGIRRE, J. M. 2013. Estilo jarraituaren alde. <http://31eskutik.com/category/jesus-maria-agirre/>. 11/03/2016.
- ALBERDI, X. and Sarasola, I. 2001. *Euskal estilo libururantz. Gramatika, estiloa eta hiztegia*. Bilbao. EHU. Euskara Institutua.
- AMURIZA, X. 2012. *Zazpi ebidentzia birjaiotzarako*. Lanku, Donostia.
- ARISTEGIETA, X. 2012. Zenbait ohar euskarazko hitz-ordenari buruz. Euskera: Euskaltzaindiaren lan eta agiriak= Trabajos y actas de la Real Academia de la Lengua Vasca = Travaux et actes de l'Academie de la Langue basque VOL. 57: 3 PAG. 505-522. Euskaltzaindia = Real Academia de la Lengua Vasca.
- ESNAL, P. 2011. *Hitz-ordena: Erabilera estrategikoa*. Euskaltzaindia. HERRIA.
- GARCÍA, J. 1996. *Joskera lantegi*. IVAP-HEAA. Vitoria-Gasteiz.
- GARZIA, J. 2014. *Esaldiaren antolaera: funtzio informatiboak gako*. EHU AZ. Bilbao.
- HIDALGO, V. 2002. Hitzen ordena esaldian. *Senez* 25, pag 75-109, EIZIE.

- IRUSKIETA, M. Maia, J. Alkorta, J. Atutxa, U. 2016. A preliminary study of the rear-burden: annotation and analysis. *XVII congreso internacional SEDLL: Educación literaria*. Badajoz.
- KALTZAKORTA, M. 2012. *Zelan ordenatu egoki informazioa testuetan-I*. Join-edition by some City Councils. Lekeitio.
- LEHMANN, C. 1985. Towards a typology of clause linkage. *Conference on Clause Combining* VOL 1 pag. 181-248.
- MAIA, J. 2014. Euskal hitz-ordenaren norabidea: idealismo ideologikotik pragmatismo komunikatiborantz. *Fontes Linguae Vasconum: Studia et Documenta* Vol. 46: 117 pag. 121-168. Institución Príncipe de Viana.
- MAIA, J. 2015. On Criteria of Professionals of the Language about the Back-burden in Basque. *Procedia* 212 pag. 67-73, Elsevier.
- MAIA, J. Larrea, K. 2013. Hitzen hurrenkera 'gogortuegia' erabiltzen ote dugu idatzian? In Alberdi, X. & Salaburu, P. (eds.). *Terminologia naturala eta terminologia planifikatua euskararen normalizazioari begira*. 85-97. UPV/EHU. Bilbao
- OTEGI, A. Imaz, O. Díaz de Ilarraza, A. Iruskieta, M. Uria, L. 2017. ANAL-HITZA: a tool to extract linguistic information from large corpora in Humanities research. *Procesamiento del Lenguaje Natural* 58: 77-84.
- RUBIO, J. 2002. *Euskararen garabideak*. Alberdania. Irun.