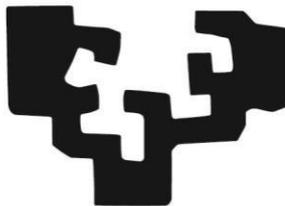


eman ta zabal zazu



Universidad
del País Vasco

Euskal Herriko
Unibertsitatea

TRABAJO DE FIN DE GRADO

**ACCESIBILIDAD EN TRADUCCIÓN E
INTERPRETACIÓN:
ANÁLISIS DE LA CALIDAD DE SUBTÍTULOS
AUTOMÁTICOS EN EUSKERA GENERADOS
CON RECONOCIMIENTO DEL HABLA**

Jon Hidalgo Chamero

Tutora: Ana Tamayo Masero

Grado en Traducción e Interpretación 2022/2023

Departamento de Filología Inglesa y Alemana y Traducción e Interpretación

Resumen

El auge de la automatización ha supuesto un gran cambio en el ámbito de la traducción y la interpretación. Además del rápido desarrollo de la traducción automática, se ha acontecido la evolución de los sistemas de reconocimiento automático del habla. Estos sistemas pueden utilizarse para garantizar la accesibilidad a los contenidos audiovisuales, por ejemplo, mediante la subtitulación para personas sordas, práctica tratada en este trabajo.

El presente Trabajo de Fin de Grado aborda la accesibilidad en la traducción e interpretación mediante un análisis cualitativo y cuantitativo de la calidad de los subtítulos en euskera, generados mediante estos sistemas. Los subtítulos automáticos analizados en este trabajo han sido generados mediante la última versión del *software* ADITU actualizada en marzo de 2023, de la Fundación Elhuyar. Se han analizado 20 muestras, de una duración de alrededor de 5 minutos cada una, que se corresponden con grabaciones realizadas en mayo de 2022 de distintas secciones de los informativos de la cadena pública vasca ETB1.

Para poder proceder con el análisis cuantitativo, se ha tomado como base el modelo NER (Romero-Fresco y Martínez Pérez, 2015), en el que se evalúa la precisión de los subtítulos mediante la detección y clasificación de los errores que afectan a la comprensión. Este modelo establece una tasa de precisión aceptable a partir del 98 % de precisión en el reconocimiento. Asimismo, en el análisis cualitativo se presentan varias cuestiones recurrentes a lo largo de las muestras que son importantes a la hora de estudiar estos subtítulos, tales como los nombres propios, la identificación del hablante y el uso de *euskalkis*, entre otros.

Por último, se ha efectuado una comparación de los resultados con un análisis previo en el que se utilizaron las mismas muestras y el mismo *software* de reconocimiento del habla en su versión de junio de 2022 (véase Tamayo y Ros-Abaurrea, en prensa). Se parte de la hipótesis de que la calidad de los subtítulos será mayor en el presente análisis.

Palabras clave: Accesibilidad, subtitulación para personas sordas, euskera, modelo NER, reconocimiento automático del habla.

Laburpena

Azken urteotan agitzen ari den automatizazioaren gorakadak aldaketa handi bat ekarri du itzulpengintza eta interpretazioaren arloan. Itzulpen automatikoaren garapen azkarrak gainera, hizketa ezagutzeko sistema automatikoen bilakaera jazo da. Sistema hauek ikus-entzunezko edukien irisgarritasuna bermatzeko erabili daitezke, esate baterako, pertsona gorrentzako azpigituluak sortuz. Lan hau praktika honi buruz arituko da.

Gradu Amaierako Lan honek irisgarritasuna jorratzen du itzulpengintza eta interpretazioaren arloan, hizketa ezagutzeko sistema automatikoen bidez sortutako euskarazko azpigituluen kalitatea aztertuz. Horretarako, analisi kualitatibo eta kuantitatibo bat egin da. Lan honetan aztertutako azpigitulu automatikoak Elhuyar Fundazioko ADITU *software*-aren azken bertsioarekin sortu dira, 2023ko martxoan eguneratua. Aztertutako 20 laginetako bakoitzak 5 minutu inguruko iraupena du eta guztiak ETB1 euskal kate publikoko albistegietako hainbat atalen grabazioei dagozkie, 2022ko maiatzean igorri zirenak.

Laginen analisi kuantitatiboa egin ahal izateko, NER eredua (Romero-Fresco eta Martínez Pérez, 2015) hartu da oinarri bezala, non azpigituluen zehaztasuna aztertzen den, haien ulermenari eragiten dieten akatsak hauteman eta sailkatuz. Eredu honen arabera, azpigituluen zehaztasun-tasa % 98tik gorakoa izan behar da hauek onargarri bezala kontuan har daitezen. Era berean, analisi kualitatiboan laginetan behin eta berriz agertu diren eta azpigituluak aztertzerakoan garrantzitsuak diren hainbat gai aurkezten dira, hala nola, izen bereziak, hizlariaren identifikazioa eta euskalkien erabilera, besteak beste.

Azkenik, analisi hau alde aurretik egindako beste analisi batekin alderatu da, lagin eta hizketa ezagutzeko sistema automatiko bera erabili zuena, 2022ko ekainaren bertsioan (ikusi Tamayo eta Ros-Abaurrea, prentsan). Halaber, lan honen analisisian azpigituluen kalitatea handiagoa izango dela dion hipotesitik abiatzen da.

Hitz gakoak: Irisgarritasuna, pertsona gorrentzako azpigituluak, euskara, NER eredua, hizketaren ezagutza automatikoa.

Índice

1. Introducción	6
2. Accesibilidad a través del subtítulado.....	7
2.1. Subtitulado en directo	7
2.1.1. Subtitulado en directo en euskera.....	8
3. Metodología.....	9
3.1. El modelo NER.....	12
3.1.1. Clasificación de errores según su gravedad.....	13
3.2. Corpus	14
4. Análisis cuantitativo.....	16
4.1. Tasa de precisión	16
4.2. Errores y correcciones.....	17
4.3. Velocidad del subtítulado	21
5. Análisis cualitativo.....	22
5.1. Identificación del hablante	22
5.2. Nombres propios	23
5.2.1. Nombres de personas.....	23
5.2.2. Topónimos	23
5.2.3. Nombres con elementos no propios del euskera	23
5.2.4. Otros nombres propios	24
5.3. Términos en otro idioma	24
5.4. Neutralización de <i>euskalkis</i>	24
5.5. Números, siglas y símbolos.....	25
5.6. Errores de continuidad	26
6. Comparación de versiones	26
6.1. Aspectos que han mejorado.....	27
6.2. Aspectos que se han mantenido.....	28
7. Conclusiones.....	29
Referencias.....	31

Índice de abreviaturas

AD	Audiodescripción
AENOR	Asociación Española de Normalización y Certificación
AR	Tasa de precisión
ASR	Reconocimiento automático del habla
CPS	Caracteres por segundo
CR	<i>Correct Recognition</i>
DNN	Redes neuronales profundas
E2E	<i>End-to-end</i>
EE	Errores de edición
EITB	Euskal Irrati Telebista
ER	Errores de reconocimiento
ER-M	Errores de reconocimiento menores
ER-SE	Errores de reconocimiento graves
ER-ST	Errores de reconocimiento estándar
GALMA	Galician Observatory for Media Accessibility
MA	Accesibilidad a los contenidos audiovisuales
NMT	Traducción automática neuronal
SPS	Subtitulación para personas sordas
TV3	Televisió de Catalunya
TVE	Televisión Española

Lista de Gráficas

Gráfica 1: AR de las muestras.....	17
Gráfica 2: Clasificación de errores según su gravedad.....	18
Gráfica 3: Clasificación de los errores de puntuación según su gravedad.....	20
Gráfica 4: Clasificación de los errores que no son de puntuación según su gravedad ...	20
Gráfica 5: Velocidad de las muestras	21

Lista de Imágenes

Imagen 1: Excel utilizado para el análisis (I)	10
Imagen 2: Excel utilizado para el análisis (II)	10
Imagen 3: Excel utilizado para el análisis (III).....	11

Lista de Tablas

Tabla 1: Resumen de las muestras	16
Tabla 2: Resumen del número de errores y CR de las muestras.....	18
Tabla 3: Ejemplo de error de reconocimiento menor (ER-M).....	19
Tabla 4: Ejemplo de error de reconocimiento estándar (ER-ST).....	19
Tabla 5: Ejemplo de error de reconocimiento grave (ER-SE)	19
Tabla 6: Ejemplo de CR.....	19
Tabla 7: Siglas reconocidas por ADITU	25
Tabla 8: Comparación de ambas versiones	26

1. Introducción

En los últimos años, se ha podido observar un auge en la automatización de la traducción. El uso de la traducción automática como parte del proceso traductológico es cada vez más común, gracias al rápido desarrollo de nuevos sistemas, como la traducción automática neuronal (NMT, por sus siglas en inglés), que genera el texto meta tomando traducciones ya existentes como base que se combinan en relación con las probabilidades estadísticas que tienen dichas oraciones de ocurrir en conjunto (Pym y Torres-Simón, 2021).

Asimismo, los sistemas de reconocimiento automático del habla (ASR, por sus siglas en inglés) también han evolucionado. Como su mismo nombre indica, estos sistemas reconocen el habla de un archivo o de un discurso y lo convierten en texto (Malik, Malik, Mehmood y Makhdoom, 2021). En la última década, el modelo híbrido de redes neuronales profundas (DNN, por sus siglas en inglés) permitió el aumento de la precisión del ASR.

Este trabajo aborda el ASR en euskera para el subtítulo en directo, un campo de estudio minoritario y de poca trayectoria académica y práctica. El objetivo principal es analizar la calidad de los subtítulos automáticos en euskera, generados mediante ASR por el *software* ADITU, de la Fundación Elhuyar, tomando como base el modelo NER (Romero-Fresco y Martínez Pérez, 2015). Se parte de la hipótesis de que la calidad de los subtítulos analizados en este trabajo, generados con la versión de marzo de 2023 de ADITU, será mayor que la de aquellos analizados previamente por Tamayo y Ros-Abaurrea (en prensa), que fueron producidos con la versión de junio de 2022 de este mismo programa.

Para ello, en la siguiente sección, se abordará el marco teórico, donde se ahondará en la accesibilidad a través del subtítulo, haciendo hincapié en el subtítulo en directo y, en particular, en el subtítulo en directo en euskera. Por las limitaciones de espacio de este trabajo, no se abordan en el marco teórico conceptos más amplios en los que se enmarca el trabajo —como traducción audiovisual o accesibilidad audiovisual— y que son ampliamente conocidos en los Estudios de Traducción e Interpretación. A continuación, en la sección 3, se explicará la metodología seguida en el trabajo y el modelo NER utilizado para el análisis (Romero-Fresco y Martínez Pérez, 2015), así como el corpus de estudio. En las secciones 4 y 5, respectivamente, se presentan los resultados del análisis cuantitativo y cualitativo. En la sección 6 se comparan los datos con un análisis previo de

las mismas muestras con una versión anterior del programa (Tamayo y Ros-Abaurrea, en prensa). Por último, en la sección 7 se ofrecen las conclusiones de este trabajo.

2. Accesibilidad a través del subtulado

Greco (2016) describe la accesibilidad a los contenidos audiovisuales (MA, por sus siglas en inglés) como un concepto que engloba todas las prácticas y técnicas que facilitan el acceso a personas que no podrían disponer de estos contenidos en su formato original. Las dos prácticas más comunes son la audiodescripción (AD) y la subtitulación para personas sordas (SPS). Esta última está más extendida en la industria. El primer programa que utilizó este tipo de subtítulos fue el documental *Quietly in Switzerland*, emitido por la BBC en 1979. En España, la SPS apareció en la década de 1990, primero en la Televisió de Catalunya (TV3) y, posteriormente, en Televisión Española (TVE). Asimismo, Euskal Irrati Telebista (EITB) empezó a ofrecer SPS en Euskadi a partir de 1999 (Díaz Cintas, 2010).

En España, la Ley 13/2022, de 7 de julio, General de Comunicación Audiovisual (BOE, 2022) dicta la cantidad de contenido que las cadenas de televisión deben emitir con SPS, concretamente, un 80 % del contenido en cadenas privadas en abierto y un 90 % en cadenas públicas. Asimismo, para garantizar un formato y características uniformes y con un mínimo de calidad en España, existe la norma UNE 153010 (AENOR, 2012), que no es obligatoria, por lo que sirve únicamente como recomendación. Además, esta solo hace alusión a la subtitulación en castellano, por lo que las demás lenguas oficiales en España no cuentan con pautas concretas que las regulen, al menos desde AENOR (Asociación Española de Normalización y Certificación).

2.1. Subtitulado en directo

La norma UNE 153010 (AENOR, 2012: 7) define el subtulado en directo como el «subtitulado de programas en directo o programas grabados que no estaban disponibles con la antelación suficiente para preparar los subtítulos antes de su emisión».

En sus orígenes, la SPS en programas en directo se realizaba mediante estenotipia, un método de escritura rápida que se realiza mediante un teclado específico. No obstante, hoy en día no es una técnica muy extendida y, además, requiere de una formación específica, lo que la convierte en un método muy costoso (Romero-Fresco, 2011).

Otro método todavía presente hoy en día es el reablado, que Romero-Fresco (2011) caracteriza como una técnica en la que un reablador reproduce el sonido original de un contenido audiovisual en directo en un *software* de ASR y, a continuación, este convierte lo expresado en subtítulos que aparecen en pantalla. Además de reproducir lo que está escuchando, el reablador incluye ciertos aspectos lingüísticos y técnicos en el proceso, como signos de puntuación o colores. Esta técnica suele utilizarse, mayormente, de manera intralingüística.

Sin embargo, con los avances tecnológicos de los últimos años, la figura del reablador se ha empezado a dejar de lado en idiomas como el español y el inglés, y es cada vez más común que los sistemas de ASR trabajen de manera autónoma. Estos sistemas están formados por tres modelos: un modelo acústico, un modelo de lenguaje y un tercer modelo con diversas denominaciones como gramática, vocabulario o diccionario (Romero-Fresco, 2011). En primer lugar, el modelo acústico está constituido por datos sobre el habla que incluyen una gran cantidad de archivos de audio junto con su transcripción exacta y la representación del archivo de audio en ondas acústicas. En segundo lugar, el modelo de lenguaje es un modelo computacional de generación de textos que calcula la probabilidad de aparición de una secuencia de palabras, según los textos que *alimentan* al programa. Por último, el tercer modelo está compuesto por un extenso listado de palabras, especificadas por los creadores del programa con anterioridad, para que el sistema elija qué terminología utilizar.

2.1.1. Subtitulado en directo en euskera

A día de hoy, el subtitulado en directo no es una realidad en los contenidos audiovisuales en euskera. Sin embargo, está siendo analizado por primera vez, y de manera sistemática, en el proyecto Qualisub¹, financiado por el Ministerio de Ciencia e Innovación (2021-2023). Uno de los objetivos principales de este proyecto es analizar la calidad de la subtitulación en directo en inglés, español, catalán, gallego y euskera. El proyecto está liderado por el grupo de investigación GALMA (Galician Observatory for Media Accessibility)².

Hoy en día, EITB no retransmite programas en directo con SPS, sino que los sistemas de ASR se utilizan únicamente como base para crear los subtítulos de programas

¹ Página web del proyecto: <https://qualisub.webs.uvigo.es/>

² Página web del grupo de investigación: <https://galmaobservatory.webs.uvigo.es/>

pregrabados (Tamayo y Ros-Abaurrea, en prensa). Hasta donde llega nuestro conocimiento, la subtítulos en directo en euskera mediante estos *softwares* nunca ha sido analizada, salvo en Tamayo y Ros-Abaurrea (en prensa).

3. Metodología

En este trabajo, se han analizado cuantitativa y cualitativamente subtítulos en euskera creados mediante un sistema de ASR. Los subtítulos analizados son intralingüísticos, según la clasificación de Díaz Cintas (2012), ya que el audio y los subtítulos comparten el mismo código lingüístico.

Para ello, se ha utilizado la versión monolingüe en euskera del *software* ADITU³, de la Fundación Elhuyar, en su versión de marzo de 2023. Este programa vio la luz por primera vez en 2020 y es capaz de transcribir y subtítular archivos de vídeo y audio, tanto pregrabados como en directo. El *software* trabaja con euskera y castellano y, en su versión actual, también funciona correctamente con archivos bilingües.

El análisis cuantitativo se ha realizado según el modelo NER (Romero-Fresco y Martínez Pérez, 2015), aunque con algunas variaciones que se comentan más abajo. Los subtítulos se han comparado con una transcripción de las muestras hecha a mano para un análisis previo (Tamayo y Ros-Abaurrea, en prensa); a excepción de la transcripción de la muestra 1, realizada expresamente para este trabajo, puesto que no había podido analizarse previamente. La clasificación de errores y el análisis se han realizado a mano en plantillas específicas de Excel⁴ utilizadas para la puesta en práctica del modelo NER, las cuales se pueden observar en las siguientes imágenes:

³ Página web del programa: <https://aditu.eus/>

⁴ Los archivos con el análisis están disponibles en:

https://drive.google.com/drive/folders/1R1ZsFpxthQj2YTDEH3tDenojuxNWDbc?usp=share_link

Tipo de segmento	Nº de subt	Tiempo inicial	Tiempo final	Duración	Caracteres	Texto del subtítulo	CPS	Nº palabras en subtítulo	Nº signos puntuación en subtítulo	Transcripción (fila 1)	Transcripción (fila 2)	Nº palabras en transcripción	Nº signos puntuación en transcripción
Presentador	1	0:00:00.43	0:00:01.08	0,65	10	HZLR2: Egi	15,38	2	1			0	0
Presentador	2	0:00:04.41	0:00:07.20	2,79	69	HZLR3: Bilboko usteko hiltzaileak gaua pasatu du Irungo Ertzaintzaren	24,73	9	1	Bilboko usteko hiltzaileak gaua pasatu du Irungo Ertzaintzaren		8	0
Presentador	3	0:00:07.21	0:00:09.23	2,02	40	komisarian, atzo bere burua eman ondoren.	19,80	6	2	komisarian, atzo bere burua eman ondoren.		6	2
Presentador	4	0:00:09.26	0:00:12.75	3,49	62	Lehen momentutik heriotzekin zerikusirik ez duela adierazi zien	17,77	8	0	Lehen momentutik heriotzekin zerikusirik ez duela adierazi zien		8	0
Presentador	5	0:00:12.83	0:00:14.92	2,09	35	ertzainei, baina arratsaldean horren	16,75	4	1	ertzainei, baina arratsaldean horren		4	1
Presentador	6	0:00:14.93	0:00:17.53	2,6	44	inguruko argibideak emango ditu Barne Sailak.	16,92	6	1	inguruko argibideak emango ditu Barne Sailak.		6	0
Presentador	7	0:00:17.83	0:00:18.66	0,83	14	Txurdinagatik.	16,87	1	1	Txurdinagatik.		1	1
Presentador	8	0:00:18.67	0:00:21.03	2,36	33	Baina bi leku horietara hurbilduko	13,98	5	0	Baina bi leku horietara hurbilduko		5	0
Presentador	9	0:00:21.04	0:00:23.26	2,22	46	gara, Irun eta Txurdinagara, azken datuen bila.	20,72	7	3	gara, Irun eta Txurdinagara, azken datuen bila.		7	3
Presentador	10	0:00:23.27	0:00:26.66	3,39	67	Bertan daude Carmina, esan diren eta Maider beste lgarbide atxilotu.	19,76	10	2	Bertan daude Garbifle San Miguel eta Maider Bestegi. Garbifle, atxilotuak		10	2
Presentador	11	0:00:26.67	0:00:28.55	1,88	33	Oraindik hor jarraitzen du, ezta?	17,55	5	2	oraindik hor jarraitzen du, ezta?		5	2
Presentador	12	0:00:30.74	0:00:34.81	4,07	33	HZLR4: Momentuz, bai, Irungo Ertzaintzaren				Ertzaintzaren komisaldegi			

Imagen 1: Excel utilizado para el análisis (I)

Problemas de precisión	Nº total de errores	Nº total de errores E	Nº total de errores R	Valor total de los errores	Error 1	Tipo de error 1	Error 2	Tipo de error 2	Error 3	Tipo de error 3	Error 4	Tipo de error 4	Error 5	Tipo de error 5	Error 6	Tipo de error 6	Error 7	Tipo de error 7	Error 8	Tipo de error 8	Error 9	Tipo de error 9	Error 10	Tipo de error 10	CR	Accuracy rate	Reduction Rate	Observaciones		
ER-ST	1,00	0,00	1,00	0,50	0,50	ER-Standard	0		0		0		0		0		0		0		0		0		0	0	0	83,33%		
HZLR2 = ER-M	1,00	0,00	1,00	0,25	0,25	ER-Minor	0		0		0		0		0		0		0		0		0		0	0	0	97,50%	100,00%	
Sailak = ER-M (puntu)	1,00	0,00	1,00	0,25	0,25	ER-Minor	0		0		0		0		0		0		0		0		0		0	0	0	96,43%	100,00%	
Carmina, esan diren = ER-ST	0,00	0,00	0,00	0,00	0		0		0		0		0		0		0		0		0		0		0	0	0	100,00%		
beste = ER-ST	0,00	0,00	0,00	0,00	0		0		0		0		0		0		0		0		0		0		0	0	0	100,00%		
lgarbide = ER-ST	0,00	0,00	0,00	0,00	0		0		0		0		0		0		0		0		0		0		0	0	0	100,00%		
atxilotu = ER-ST	4,00	0,00	4,00	2,00	0,50	ER-Standard	0,50	ER-Standard	0,50	ER-Standard	0,50	ER-Standard	0		0		0		0		0		0		0	0	0	83,33%		
HZLR4 = ER-M	1,00	0,00	1,00	0,25	0,25	ER-Minor	0		0		0		0		0		0		0		0		0		0	0	0	97,50%	100,00%	Más texto en pantalla
	0,00	0,00	0,00	0,00	0		0		0		0		0		0		0		0		0		0		0	0	0	100,00%		
	0,00	0,00	0,00	0,00	0		0		0		0		0		0		0		0		0		0		0	0	0	100,00%	Más texto en pantalla	

Imagen 2: Excel utilizado para el análisis (II)

En la Imagen 1 e Imagen 2 se puede observar la pestaña principal del Excel. En ella, se presentan todos los datos relevantes de las muestras, como el texto del subtítulo, la transcripción o los tipos de errores. Se ha utilizado un Excel para cada muestra.

En cuanto a las variaciones del modelo NER, en este TFG solo se han tenido en cuenta los errores de reconocimiento (ER) y se han dejado de lado los errores de edición (EE), debido a que los subtítulos han sido creados mediante un sistema de ASR y no han sido editados. Asimismo, mientras que el modelo NER tampoco tiene en cuenta las etiquetas de identificación de personajes, en este trabajo sí se han valorado, ya que el programa ADITU, a diferencia de otros sistemas de ASR, sí permite la identificación de personajes.

No obstante, no se ha analizado si el formato de la etiqueta en sí es el adecuado siguiendo la norma UNE 153010 (AENOR, 2012) u otras guías, sino si la adición de etiquetas para la identificación de personajes conlleva, o no, errores según el modelo NER.

En este trabajo, también se ha analizado la velocidad de los subtítulos (Imagen 3), definida como la velocidad en la que un subtítulo se presenta en pantalla, y que se calcula dividiendo el número de caracteres del subtítulo con el tiempo en el que son visibles en pantalla (Fresno y Sepielak, 2022).

Código	Tipo de segmento	Nº de subt	Tiempo inicial	Tiempo final	Duración	Caracteres	Texto del subtítulo	CPS	Nº palabras en subtítulo	WPM	Canal
ETB_7	Presentador	1	0:00:00.43	0:00:01.08	0,65	10 HZLR2: Egu	15,38	2	184,62		
ETB_7	Presentador	2	0:00:04.41	0:00:07.20	2,79	HZLR3: Bilboko ustezko hitzaileak gaua pasatu du Irungo	24,71	9	193,55		
ETB_7	Presentador	3	0:00:07.31	0:00:09.23	2,02	49 Ertaintzaren	19,80	6	178,22		
ETB_7	Presentador	4	0:00:09.26	0:00:12.75	3,49	Lehen momentutik heriotzekin zerikusirik ez duela adierazi	17,77	8	137,54		
ETB_7	Presentador	5	0:00:12.83	0:00:14.32	2,09	35 ertzainei, baina arnastaldian horren	16,75	4	114,83		
ETB_7	Presentador	6	0:00:14.93	0:00:17.53	2,6	44 Inguurako argibideak emango ditu Barne Sailak.	16,82	6	138,46		
ETB_7	Presentador	7	0:00:17.83	0:00:18.66	0,83	14 Txurdinagatik.	16,87	1	72,29		
ETB_7	Presentador	8	0:00:18.67	0:00:21.03	2,36	33 Baina bi leku horietara hurbilduko	13,98	5	127,12		
ETB_7	Presentador	9	0:00:21.04	0:00:23.26	2,22	46 gara, Inui eta Txurdinagata. izanen datuen bila.	20,72	7	189,19		
ETB_7	Presentador	10	0:00:23.27	0:00:26.66	3,39	Berlan daude Carmina, esan diren eta Maider beste igaribide	15,76	10	176,99		
ETB_7	Presentador	11	0:00:26.67	0:00:28.55	1,88	67 atxilotu.	17,55	5	159,57		
ETB_7	Presentador	12	0:00:30.74	0:00:34.81	4,07	33 Oramdik hor jarraitzen du, eta? HZLR3: Momentuz, bai, Irungo Ertaintzaren komisaldigi	14,99	7	103,19		
ETB_7	Reportero	13	0:00:34.82	0:00:37.32	2,5	61 honetan	12,40	4	96,00		
ETB_7	Reportero	14	0:00:37.35	0:00:41.27	3,92	31 dago Bilboko hilketen susmagarri	11,99	6	91,84		Medio CPS 16,69
ETB_7	Reportero	15	0:00:41.38	0:00:47.18	5,6	47 nagusia, epailearen aurrean noiz deklaratu zain.	11,96	10	107,14		
ETB_7	Reportero	16	0:00:47.19	0:00:49.08	1,89	Egia da? 72 ordu egon dutekeela egon dutekeela komisaldigi	15,05	4	126,98		
ETB_7	Reportero	17	0:00:49.12	0:00:52.10	2,98	67 honetan	16,78	8	162,07		
ETB_7	Reportero	18	0:00:53.11	0:00:53.87	0,76	36 epailearen aurrean deklaratu ondoren,	11,84	1	78,99		
ETB_7	Reportero	19	0:00:53.16	0:00:58.40	5,24	50 eta, gainera, epaileak epe hori lutzatu egin dezake.	12,02	8	91,80		
ETB_7	Reportero	20	0:00:58.41	0:01:00.41	2,0	9 Momentuz.					
ETB_7	Reportero	21	0:01:00.41	0:01:03.41	3,0	Bilbon habe urrian gertatutako gizonezko baten heriotzagatik					
ETB_7	Reportero	22	0:01:03.41	0:01:05.41	2,0	63 eta					
ETB_7	Reportero	23	0:01:05.41	0:01:07.41	2,0	abenduaren egindako hilketak saiaera batengatik ikertzen ari					

Imagen 3: Excel utilizado para el análisis (III)

En las últimas décadas se han publicado artículos en los que se estudia la velocidad de los subtítulos y se ha hablado de la capacidad de lectura de las personas sordas y el esfuerzo cognitivo (véase, entre muchos otros, de Linde y Kay, 1999, y Díaz Cintas y Remael, 2007), la legibilidad de los subtítulos (véase Díaz Cintas, 2003, e Ivarsson y Carroll, 1998) y la reducción de información (véase Romero-Fresco, 2009). El análisis de la velocidad en este TFG tiene en cuenta publicaciones previas relacionadas con la velocidad del subtítulo para personas sordas.

Es importante subrayar que el análisis cuantitativo se ha realizado tanto teniendo en cuenta los errores de puntuación como excluyéndolos, por diferentes motivos. En primer lugar, porque se ha demostrado que es el ámbito en el que los sistemas de ASR suelen cometer más errores y, en segundo lugar, porque no todo lo generado mediante estos programas tiene en cuenta la puntuación (Hlubík, Španěl, Boháč y Weingartová, 2020).

Además del análisis cuantitativo, también se ha realizado un análisis cualitativo. Gracias a este, se han detectado a lo largo de las muestras varias cuestiones recurrentes que son importantes a la hora de estudiar los subtítulos generados mediante programas de ASR. En las secciones 4 y 5, respectivamente, pueden consultarse los resultados del análisis cuantitativo y cualitativo.

3.1. El modelo NER

El objetivo del modelo NER (Romero-Fresco y Martínez Pérez, 2015) es evaluar la precisión de los subtítulos en directo mediante la detección de errores que afectan a la comprensión de estos y, en consecuencia, dificultan el entendimiento del receptor.

Para la creación del modelo NER, se utilizó como base el modelo presentado en Romero-Fresco (2011), originalmente denominado «modelo NERD», adaptándolo a las características y necesidades de la subtitulación en directo. No obstante, en la industria es mucho más frecuente el uso del modelo WER (Romero-Fresco y Martínez Pérez, 2015). El problema del modelo WER es que etiqueta como error todo aquello que no se corresponde exactamente con el audio (por ejemplo, que el programa subtitule «*can't*» cuando el hablante ha dicho «*cannot*»). Por lo tanto, el modelo WER no tiene en cuenta al receptor y se centra, únicamente, en la correspondencia audio-subtítulo. El modelo NER surgió como alternativa al WER y se centra en marcar los errores de reconocimiento (ER) y de edición (EE) teniendo la comprensión del subtítulo como punto de partida. Así, mientras el modelo WER es fácilmente automatizable, el modelo NER precisa de un humano (de momento) para analizar los errores y la gravedad de los mismos.

El modelo NER calcula la precisión de los subtítulos mediante la siguiente fórmula:

$$Accuracy\ rate\ (AR) = \frac{N - E - R}{N} \times 100$$

Los valores que aparecen en la fórmula son los siguientes:

- N: número de palabras en el audio, incluyendo también cualquier tipo de comando, como signos de puntuación, entre otros.
- E: errores de edición, normalmente relacionados con la puntuación o el uso incorrecto de las mayúsculas, entre otros. Los errores de edición se clasifican en errores menores, estándar y graves, mediante la comparación del texto original y el texto expresado en los subtítulos, y se penalizan con -0.25, -0.5 y -1 puntos, respectivamente.

- R: errores de reconocimiento, provocados por errores de pronunciación y de escucha o causados por la tecnología utilizada para producir los subtítulos; como inserciones, omisiones o sustituciones. Los errores de reconocimiento se clasifican en errores menores, estándar y graves, mediante la comparación del texto original y el texto de los subtítulos, y se penalizan con -0.25, -0.5 y -1 puntos, respectivamente.

Para el análisis de este trabajo, solo se han tenido en cuenta errores de reconocimiento (ER) y no errores de edición (EE), puesto que estos últimos solo se dan cuando los subtítulos se pueden editar antes de enviarlos (en rehablado, por ejemplo) y no es el caso de los que han sido analizados en este TFG.

En el modelo NER, y en este trabajo, también se tienen en cuenta las correcciones automáticas realizadas por el programa cuando este detecta redundancias o repeticiones que no le restan comprensión al contenido audiovisual. Estas correcciones se denominan CR (*Correct Recognition*, en inglés).

Para que la precisión de un subtítulo pueda considerarse aceptable, deberá alcanzar una AR superior al 98 %, de acuerdo con lo establecido en Romero-Fresco (2016) y Romero-Fresco y Pöchhacker (2017).

Es importante recalcar que, en este trabajo, con el fin de diferenciar la precisión del reconocimiento del habla de otros factores, se consideran errores de puntuación todos aquellos que no tienen que ver con el reconocimiento del habla propiamente dicho (como errores de tildes, identificación de personajes u ortotipográficos).

3.1.1. Clasificación de errores según su gravedad

Tal y como expresan Romero-Fresco y Martínez Pérez (2015), los errores pueden clasificarse en tres grupos, teniendo en cuenta su gravedad.

Los errores menores son aquellos que el receptor puede detectar como error y no impiden la correcta comprensión de los subtítulos. Estos errores se puntúan con -0.25 puntos.

Los errores estándar se dan cuando se omite o se altera una unidad de información. El espectador puede detectarlos como error, pero impiden la correcta comprensión del subtítulo. Estos errores se puntúan con -0.5 puntos.

Los errores graves son aquellos que cambian el significado del texto original y, en consecuencia, los fragmentos alterados obtienen un nuevo significado que podría tener

sentido en un contexto en concreto. Los receptores no suelen percibirlos como errores, debido a este sentido nuevo, por lo que puede crearse un proceso de desinformación. Estos errores se puntúan con -1 punto.

3.2. Corpus

Para este trabajo, se ha llevado a cabo un análisis, tanto cuantitativo como cualitativo, de 101.67 minutos (1:41:40 horas) y 1973 subtítulos en euskera, creados automáticamente por el programa ADITU, de la Fundación Elhuyar, en su versión de marzo de 2023. Se han analizado un total de 20 muestras⁵ pertenecientes a informativos en euskera de la cadena de televisión pública vasca ETB1, cada una con una duración de alrededor de 5 minutos, emitidos entre los días 2 y 15 de mayo de 2022 a distintas horas del día. Las muestras se corresponden con distintas partes del telediario (deportes, el tiempo, entrevistas...), por lo que son diversas. Cabe destacar que estas muestras no son siempre monolingües, pero que solo se han analizado aquellos subtítulos en los que el audio sea totalmente en euskera, obviando los subtítulos en castellano, inglés u otros idiomas. Las muestras se han recogido a través de la tutora de este trabajo, Ana Tamayo Masero, puesto que ya se habían utilizado para realizar un análisis previo (Tamayo y Ros-Abaurrea, en prensa).

La información correspondiente a los archivos analizados puede observarse en la Tabla 1:

ID	Día y horario	TCR inicio	TCR final	Tipo	Min.	N.º sub.	Notas
ETB_1	02/05/2022 9-11 horas	0:00:10	0:05:05	Inicio, noticias, presentadores, gente de la calle	0:04:55	78	Muestra en más de un idioma: EU, ES.
ETB_2	02/05/2022 9-11 horas	0:39:58	0:45:00	Entrevista	0:05:02	102	Muestra en más de un idioma: EU, ES.
ETB_3	03/05/2022 14-15 horas	0:20:37	0:25:38	Noticias, presentadores, reporteros	0:05:01	89	Muestra en más de un idioma: EU, ES.
ETB_4	03/05/2022 14-15 horas	0:54:30	0:59:32	El tiempo	0:05:02	129	

⁵ Las muestras pueden encontrarse en el siguiente enlace:
https://drive.google.com/drive/folders/1cYnpc2zhWgKZ5UGc1beq2RbwgquWG3F7?usp=share_link

ETB_5	04/05/2022 20-21 horas	0:10:09	0:15:08	Noticias, presentadores, reporteros, gente de la calle	0:04:59	89	Muestra en más de un idioma: EU, ES.
ETB_6	04/05/2022 20-21 horas	0:38:14	0:43:29	Deportes	0:05:15	103	Muestra en más de un idioma: EU, ES.
ETB_7	06/05/2022 14-15 horas	0:07:11	0:12:18	Primera noticia, presentador, reporteras	0:05:07	95	Muestra en más de un idioma: EU, ES.
ETB_8	06/05/2022 14-15 horas	0:43:47	0:49:02	Deportes	0:05:15	102	Muestra en más de un idioma: EU, ES.
ETB_9	07/05/2022 20-21 horas	0:31:35	0:36:47	Últimas noticias (cultura, historia), presentadora, reporteras	0:05:12	99	Muestra en más de un idioma: EU, ES.
ETB_10	07/05/2022 20-21 horas	0:53:10	0:58:30	El tiempo	0:05:20	110	
ETB_11	11/05/2022 9-11 horas	0:00:17	0:05:20	Noticias, presentadores	0:05:03	85	Muestra en más de un idioma: EU, ES, CA.
ETB_12	11/05/2022 9-11 horas	0:27:35	0:32:30	Debate y entrevista en plató	0:04:55	86	Muestra en más de un idioma: EU, ES.
ETB_13	12/05/2022 14-15 horas	0:24:10	0:29:14	Noticias	0:05:04	83	Muestra en más de un idioma: EU, ES.
ETB_14	12/05/2022 14-15 horas	0:33:19	0:38:20	Últimas noticias (cultura, historia), presentador, gente de la calle	0:05:01	104	Acento y variedad de Iparralde
ETB_15	13/05/2022 20-21 horas	0:00:10	0:05:17	Titulares con reporteros	0:05:07	89	Muestra en más de un idioma: EU, ES.

ETB_16	13/05/2022 20-21 horas	0:51:20	0:56:29	El tiempo	0:05:09	134	
ETB_17	14/05/2022 20-21 horas	0:09:35	0:14:35	Noticias	0:05:00	87	Muestra en más de un idioma: EU, EN, DE, UK.
ETB_18	14/05/2022 20-21 horas	0:42:58	0:48:05	Deportes	0:05:07	112	Muestra en más de un idioma: EU, EN.
ETB_19	15/05/2022 9:10-9:40 horas	0:25:10	0:30:00	<i>Iparraldearen orena</i>	0:04:50	94	Acento y variedad de Iparralde
ETB_20	15/05/2022 9:10-9:40 horas	0:04:48	0:10:04	<i>Iparraldearen orena</i>	0:05:16	103	Acento y variedad de Iparralde Muestra en más de un idioma: EU, FR.
TOTAL					101.67 min.	1973 sub.	

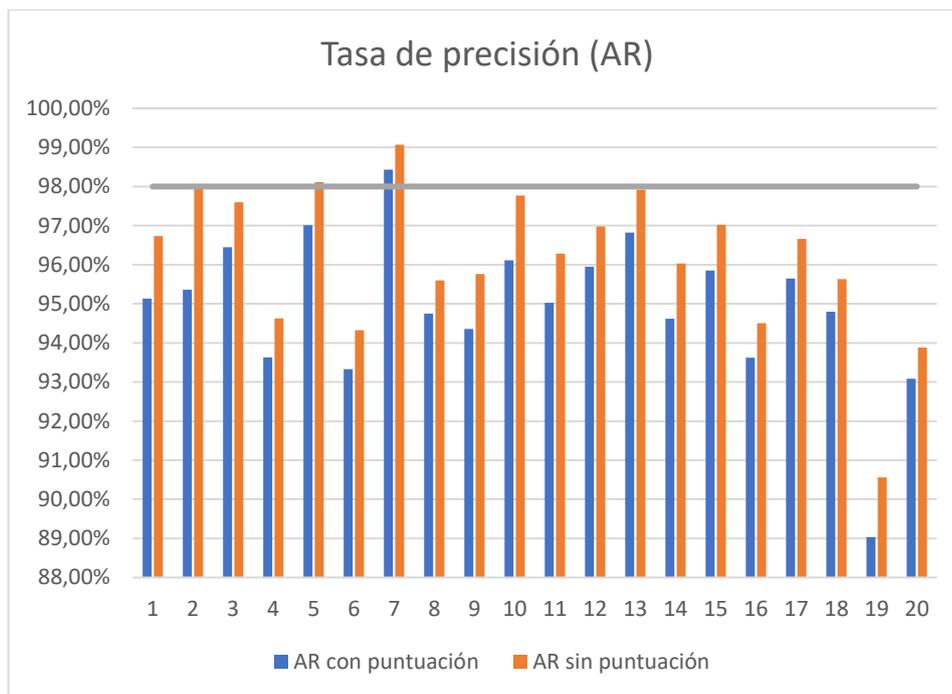
Tabla 1: Resumen de las muestras

4. Análisis cuantitativo

A continuación, se analizarán estos subtítulos de manera cuantitativa, tomando como base el modelo NER. Por un lado, se observará la AR de los subtítulos tanto teniendo en cuenta los errores relacionados con la puntuación como excluyéndolos. Tal y como se expresa en Romero-Fresco (2016) y Romero-Fresco y Pöchhacker (2017), el modelo NER establece una precisión aceptable a partir de una AR del 98 %. Por otro lado, se analizarán los errores que se han encontrado, clasificándolos según su tipo y gravedad. Por último, se observará la velocidad de los subtítulos analizados.

4.1. Tasa de precisión

En la Gráfica 1 se puede observar la AR de cada una de las muestras, tanto contando como sin contar los errores de puntuación. Teniendo en cuenta todos los errores, la media de la AR en el reconocimiento del habla para el subtítulo en directo en euskera de los archivos analizados es del 94.95 %. Si no se tienen en cuenta los errores de puntuación, la tasa media es del 96.15 %. Por lo tanto, los subtítulos no superan, de media, el mínimo aceptable del 98 % (marcado con una línea gris en la Gráfica 1), según el modelo NER.



Gráfica 1: AR de las muestras

Si examinamos las muestras por separado, la muestra 7 es la única que, siguiendo las tasas establecidas por el modelo NER, podría considerarse aceptable tanto teniendo en cuenta los errores de puntuación (98.43 %) como sin ellos (99.07 %). En la muestra 7, el presentador y las reporteras utilizan el euskera *batua*, es decir, un euskera estándar y unificado, sin dialectos. Además, el contenido narrado no es espontáneo, está guionizado. Asimismo, la muestra 5 también sobrepasa el umbral del 98 %, pero solo si no se tienen en cuenta los errores de puntuación (98.11 %).

Cabe destacar la AR de la muestra 19, que es la más baja de todas, con un 89.03 % contando los errores de puntuación y un 90.56 % sin ellos. Esta muestra, junto con la muestra 20, pertenece a la sección del telediario dedicada a las noticias de Iparralde denominada *Iparraldearen orena* y presenta una entrevista a un granjero de esta zona. Por consiguiente, el euskera utilizado no es el *batua*, sino la variante de Iparralde, que cuenta con un acento francés, más difícil de identificar para el programa.

En cuanto a las demás muestras, su AR oscila entre el 93.08 % y el 97.01 %, una tasa no aceptable según lo estipulado en el modelo NER.

4.2. Errores y correcciones

En las 20 muestras, se han observado un total de 1975 errores, de los cuales 1089 son menores, 802 estándar y 84 graves, y 76 CR, tal y como se muestra en la Tabla 2:

ID	Errores totales	ER-M	ER-ST	ER-SE	CR
ETB 1	81	53	24	4	6
ETB 2	104	70	30	4	7
ETB 3	68	42	22	4	0
ETB 4	158	69	82	7	12
ETB 5	60	42	15	3	6
ETB 6	126	63	60	3	6
ETB 7	35	26	8	1	4
ETB 8	114	68	45	1	0
ETB 9	107	62	39	6	0
ETB 10	83	53	25	5	6
ETB 11	83	49	29	5	0
ETB 12	75	48	25	2	4
ETB 13	58	41	17	0	1
ETB 14	108	54	51	3	3
ETB 15	78	59	16	3	1
ETB 16	155	83	59	13	9
ETB 17	70	36	30	4	0
ETB 18	115	68	44	3	7
ETB 19	164	43	113	8	0
ETB 20	133	60	68	5	4
TOTAL	1975	1089	802	84	76

Tabla 2: Resumen del número de errores y CR de las muestras

En cuanto a la gravedad de los errores, los porcentajes son los siguientes:



Gráfica 2: Clasificación de errores según su gravedad

Los errores de reconocimiento menores (ER-M) no dificultan la comprensión del subtítulo, por lo que el usuario puede entender el contenido audiovisual correctamente. Por ejemplo, el siguiente error se considera menor:

Muestra	Transcripción	Subtítulo
ETB_5	diru iturria agortuko zaio Moskuri. Baina,	diru iturri agortuko zaio Moskuri, baina,

Tabla 3: Ejemplo de error de reconocimiento menor (ER-M)

Los errores de reconocimiento estándar (ER-ST) son aquellos que impiden que el usuario comprenda correctamente el subtítulo, pero que pueden detectarse como error. Por ejemplo, el siguiente error se considera estándar:

Muestra	Transcripción	Subtítulo
ETB_6	azterketa berezia, mimo bereziarekin prestatzen duguna.	Azterketa berezia, mimo berezia egin , prestatzen duguna.

Tabla 4: Ejemplo de error de reconocimiento estándar (ER-ST)

Los errores de reconocimiento graves (ER-SE) no son identificables sin tener acceso al audio, por lo que los usuarios pueden no darse cuenta del error y entender algo distinto a lo expresado en el contenido audiovisual. Por ejemplo, el siguiente error se considera grave:

Muestra	Transcripción	Subtítulo
ETB_16	leku guztietatik agiri, neure sorterria, eta bertan izan da Aitor Agirrezabal.	Leku guztietatik agiri gure sorterria eta bertan izan da Aitor Agirrezabal.

Tabla 5: Ejemplo de error de reconocimiento grave (ER-SE)

Además de los errores, también se han considerado las correcciones automáticas que el programa realiza gracias a su modelo de lenguaje. Estas correcciones suelen hacerse en las marcas de oralidad y las repeticiones. Por ejemplo, la siguiente omisión se considera CR:

Muestra	Transcripción	Subtítulo
ETB_4	eta, hori , temperatura Santikutz egunekoa,	Eta temperatura Santikutz egunekoa

Tabla 6: Ejemplo de CR

Los errores de puntuación corresponden a un 31.8 % (628 errores) del total. Se han considerado errores de puntuación tanto aquellos relacionados con la puntuación en sí (falta de puntuación a final de oración, signos de puntuación sobrantes, puntos que se han reconocido como comas...), como otros no relacionados con el reconocimiento del habla (falta de tildes⁶, uso incorrecto de las mayúsculas y minúsculas, identificación de

⁶ La guía de estilo de EITB (Martín Sabarís, Cantalapiedra, Zalbidea Bengoa y Aranburu, 2016) estipula que los nombres propios deben escribirse sin marcas diacríticas en contenidos audiovisuales en euskera. Por lo tanto, los errores de tildes se han categorizado como errores de puntuación.

personajes, uso de «k» en vez de «c» en algunos nombres propios...). Para distinguir los errores de reconocimiento de otros tipos de errores (aquí denominados «de puntuación»), se ofrecen las siguientes gráficas. La Gráfica 3 muestra la gravedad de los errores de puntuación (n=628) y la Gráfica 4 muestra la gravedad de los errores de reconocimiento (n=1347).



Gráfica 3: Clasificación de los errores de puntuación según su gravedad



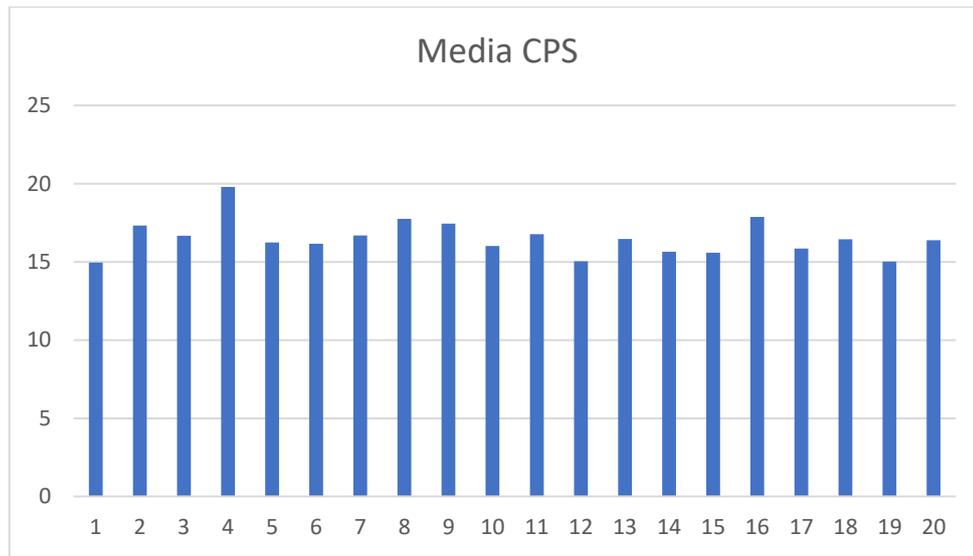
Gráfica 4: Clasificación de los errores que no son de puntuación según su gravedad

Tal y como expresa la Gráfica 3, los errores de puntuación son, en mayoría, errores menores, por lo que no impiden la comprensión del usuario. Por otro lado, en los errores que no son de puntuación, más de la mitad son errores estándar, como se puede observar en la Gráfica 4. Cabe destacar que la cantidad de errores estándar en euskera es mayor

que en otros idiomas como como el inglés (Romero-Fresco y Fresno, 2023), el gallego (Rico, Docío y García, 2020) y el catalán (Tamayo y de Higes, 2023).

4.3. Velocidad del subtítulo

Fresno y Sepielak (2022) expresan que existen ciertas discrepancias metodológicas a la hora de calcular la velocidad de los subtítulos. En este trabajo, la velocidad se ha medido mediante caracteres por segundo (CPS), como puede observarse en la Gráfica 5:



Gráfica 5: Velocidad de las muestras

Según dicta la norma UNE 153010 (AENOR, 2012), el máximo de CPS recomendado para la SPS en español es de 15 CPS. Observando la media, la única muestra que no supera esta recomendación es la muestra 1, con 14.96 CPS. Asimismo, otras muestras se encuentran próximas a los 15 CPS, aunque superan esta media: la muestra 12, con 15.03 CPS; la muestra 14, con 15.65 CPS; la muestra 15, con 15.59 CPS; la muestra 17, con 15.85 CPS; y la muestra 19, con 15.01 CPS. Cabe destacar que en la mayoría de las muestras se puede encontrar algún subtítulo a una velocidad excesiva, como el subtítulo 46 de la muestra 4 (25.6 CPS) o el subtítulo 66 de la muestra 16 (27.14 CPS).

Sin embargo, es importante recalcar que en España no existe una norma que delimite el número máximo de caracteres por segundo en euskera, por lo que la media recomendada podría no ser la misma que la utilizada en los subtítulos en castellano. A falta de un estudio de recepción y de un análisis pormenorizado de la velocidad del subtítulo, puede considerarse que la velocidad de los subtítulos de las muestras es adecuada.

5. Análisis cualitativo

Son diversos los factores que pueden influir en la calidad de los subtítulos creados mediante sistemas de ASR y, en consecuencia, en la AR de estos. Por ejemplo, como bien se ha podido observar analizando los errores de las muestras, el reconocimiento es mejor si el euskera que se utiliza en el audio es estándar. Asimismo, otros aspectos que contribuyen a la calidad de los subtítulos son el registro utilizado, el ruido de fondo y el número de hablantes por muestra, entre otros. En este apartado, se han seleccionado varios elementos recurrentes a lo largo de las muestras que merecen ser destacados.

5.1. Identificación del hablante

El *software* ADITU identifica a los distintos hablantes mediante la etiqueta «HZLR» (abreviatura de «*hizlari*», «hablante» en euskera), seguida del número que le corresponde a la persona y dos puntos. La identificación de personajes conlleva errores de todo tipo. Por un lado, a veces la etiqueta de identificación es innecesaria, puesto que el hablante aparece en pantalla y el usuario puede reconocer quién está hablando gracias a la información del canal visual. En estos casos, el error es menor (ER-M), ya que no dificulta el entendimiento del contenido audiovisual. Por otro lado, la etiqueta puede aparecer cuando no debe, por ejemplo, a mitad de intervención de un hablante. En estos casos, el error es estándar (ER-ST), puesto que puede generar confusión. En otras ocasiones, cuando se está retransmitiendo un vídeo pregrabado y el reportero es distinto, a veces no aparece una nueva etiqueta y, por este motivo, el espectador sordo puede pensar que el hablante no ha cambiado. Además, el programa a veces reutiliza la misma etiqueta numérica para marcar dos hablantes distintos, por lo que también puede llevar a confusión. En estos dos casos, el error es grave (ER-SE).

Teniendo en cuenta estos errores, podría pensarse que el programa no reconoce un nuevo hablante hasta que este no aparece en pantalla. Sin embargo, Igor Leturia, responsable de tecnologías del habla de Orai NLP Teknologiak que trabaja en el desarrollo de ADITU, expresa que la identificación de personajes se predice mediante la información acústica y los silencios entre intervenciones (Leturia, comunicación personal, en Tamayo y Ros-Abaurrea, en prensa).

5.2. Nombres propios

Uno de los errores más frecuentes en las muestras se debe al reconocimiento de los nombres propios. La mayoría de estos errores son errores estándar (ER-ST), exceptuando aquellos que tienen que ver con tildes u otros elementos propios del castellano.

Leturia (comunicación personal, en Tamayo y Ros-Abaurrea, en prensa) expresa que los errores relativos a los nombres propios no son fáciles de corregir, debido a que habría que actualizar el modelo de lenguaje de ADITU con frecuencia para poder reconocer los nombres correspondientes a hechos actuales; proceso que no es posible teniendo en cuenta el estado de desarrollo del programa. Asimismo, no solo habría que añadir los términos *per se*, sino también todas las posibles declinaciones de estos en euskera.

5.2.1. Nombres de personas

En primer lugar, el *software* no reconoce correctamente todos los nombres de personas. En un principio, podría pensarse que el problema aparece cuando se trata de nombres extranjeros o poco comunes en euskera, como Caleb Ewan (muestra 6) o Juliette Binoche (muestra 15), subtítulados como «*Kaletik joanda*» y «*Zuriak Binos*», respectivamente. No obstante, estos errores también ocurren cuando los nombres provienen del euskera, por ejemplo, Iban (reconocido como «*Egiman*» en la muestra 2) o Kepa (reconocido como «*Bake*» en la muestra 4). Estos errores se han marcado como errores estándar (ER-ST).

5.2.2. Topónimos

En el caso de los topónimos también se dan errores tanto en nombres en euskera como extranjeros. Como ejemplo de nombres provenientes del euskera tenemos «Estatu Batuetan», «Nafarroako» o «Pirineo», subtítulados incorrectamente como «*Estatu eta Batuetan*» (muestra 3), «*akropoliko*» (muestra 4) y «*Pipino*» (muestra 4), respectivamente. En cuanto a topónimos extranjeros, nos encontramos con Leipzig o Kharkiv (subtítulados como «*Leiziek*» y «*Jarkik*» en las muestras 8 y 17, respectivamente).

5.2.3. Nombres con elementos no propios del euskera

Debido a que se ha utilizado la versión monolingüe en euskera de ADITU, el programa ha creado errores al modificar y *convertir* al euskera todos aquellos elementos lingüísticos que no existen en este idioma, como las tildes y la letra «v». Por ejemplo, en la muestra 1

los apellidos de Pedro Sánchez y Yolanda Díaz se han subtitulado sin sus respectivas tildes⁷ y, en la muestra 8, el programa ha reconocido Sevilla como «*Sebilla*».

5.2.4. Otros nombres propios

El programa también ha mostrado problemas a la hora de reconocer otros nombres relacionados con acontecimientos políticos actuales, como el caso Roe versus Wade, que el *software* ha subtitulado como «*harro zuei*» (muestra 3). Asimismo, también se dan problemas con nombres de marcas comerciales, como en la muestra 13, donde el modelo de coche Renault Megane aparece como «*trenor megan*».

5.3. Términos en otro idioma

Como se puede observar en la Tabla 1, no todas las muestras son monolingües. Estos segmentos en otros idiomas no se han tenido en cuenta en el análisis, puesto que este trabajo se centra únicamente en la situación del reconocimiento del habla en euskera. No obstante, al igual que en todos los idiomas, en euskera también se utilizan préstamos y, a veces, el programa no los reconoce correctamente. En estos casos, se ha observado que pueden ocurrir dos cosas: que el *software* no reconozca el término y subtitule algo sin sentido, como es el caso de «*gigabyte*» en la muestra 11, que el programa ha reconocido como «*txikia baita*», o que el programa encuentre ciertas combinaciones de palabras en su modelo de lenguaje y utilice las reglas del euskera para redactarlo, como ocurre en la muestra 18 con «*play off-etan*», convertido en «*plai ofetan*».

De la misma manera, hoy en día es común que en el ámbito del deporte se utilicen términos en inglés. Este problema se da, por ejemplo, al hablar del campeonato mundial de ciclismo. El programa no reconoce este término («*World Tour-ekoak*») y en el subtítulo aparece «*Holt Hurrekoak*» (muestra 6). En cuanto a los nombres de los equipos, el programa reconoce el Sopela Women's Team como «*Sopela Women Eskuinez*» (muestra 6).

5.4. Neutralización de *euskalkis*

El programa reconoce mejor lo que se está diciendo si el hablante está utilizando euskera *batua* y también estandariza todo lo que se dice, neutralizando así cualquier variante del euskera. Así, «*jeitsi*» se convierte en «*jaitsi*» y «*alkatia*» en «*alkatea*», entre otros. Asimismo, en la muestra 9, el reportero lee una cita que aparece en pantalla y los

⁷ Las tildes se han contabilizado como errores en este análisis, pero la guía de estilo de EITB (Martín Sabarís, Cantalapiedra, Zalbidea Bengoa y Aranburu, 2016) no considera esto como error.

subtítulos eliminan cualquier rastro de la variante original. Este cambio también ocurre cuando la gente a la que entrevistan por la calle no utiliza el euskera *batua*.

Es cierto que estas correcciones pueden considerarse positivas, debido a que ayudan a que un público más amplio pueda comprender de manera más fácil lo que se está diciendo. No obstante, mediante esta estandarización se elimina cualquier tipo de matiz e información que la variante nos puede ofrecer, como la procedencia del hablante, entre otros factores. Además, los usuarios pueden sentir que no se les está proveyendo con un contenido exacto o *verbatim* y que los subtítulos no son fieles a lo que se dice en el audio. Esta cuestión genera un debate sobre si debería prevalecer el hecho de crear un contenido *entendible*, mediante esta estandarización del lenguaje, aunque este acto pueda no considerarse del todo *accesible*, al eliminar ciertos aspectos del contenido audiovisual original. Esta cuestión de comprensibilidad frente a accesibilidad (y sus variantes tales como subtítulos editados frente a subtítulos *verbatim*) sigue generando gran debate en encuentros académicos y de la industria, también en subtítulo en directo (véase Eugeni, 2023).

5.5. Números, siglas y símbolos

En relación con los números, el programa reconoce correctamente aquellos que hacen referencia a horas. Por ejemplo, en la muestra 2 se dice que son las diez menos cuarto («*hamarrak hogei gutxi*», en euskera) y en los subtítulos aparece «09:40». En la muestra 7 también podemos encontrar otro ejemplo, donde el programa reconoce directamente que las cuatro de la tarde («*arratsaldeko laurak*», en euskera), son las «16:00». No obstante, con las fechas se dan errores menores (ER-M), ya que el programa escribe todos los números a partir de mil con un punto. Por ejemplo, en la muestra 6, se hace mención al año 2023 y en el subtítulo aparece «2.023».

Asimismo, aunque el hablante diga el término completo, el programa es capaz de reconocer las siglas correspondientes. En la siguiente tabla pueden observarse algunas siglas que ADITU ha subtulado:

Muestra	Término	Sigla reconocida
ETB_1	Comisiones Obreras	CCOO
ETB_2	Euskal Telebista	ETB
ETB_5	Euskal Autonomia Erkidegoa	EAE
ETB_5	Europar Batasuna	EB
ETB_9	Emakume Abertzale Batza	EAB

Tabla 7: Siglas reconocidas por ADITU

Además, el programa también reconoce los símbolos correspondientes a litro («l»), grado («°») o euro («€») y los utiliza en los subtítulos en vez de escribirlos con letras.

5.6. Errores de continuidad

Este *software* de ASR a veces no es constante con sus decisiones. Por ejemplo, en la muestra 5, se listan varios países y mientras que los dos primeros sí que se escriben con mayúscula inicial, el tercero no («*Hungariak, Bulgariak, txekiar errepublikak*»). Asimismo, se han detectado errores de reconocimiento (ER) que se dan algunas veces pero no otras, incluso dentro de la misma muestra. Estos errores pueden deberse a múltiples factores como la calidad del audio o la manera en la que se han pronunciado exactamente estos términos. Como ejemplo de esto, en la muestra 6 podemos observar que «*ETBlen*» se ha subtulado como «*epe batean*», mientras que en esta misma muestra sí que se ha reconocido «*ETB*» correctamente. En la muestra 17 también ocurre lo mismo, donde «*Ukrainako*» se subtitula como «*bukaerako*», a pesar de que el nombre del país ya ha aparecido anteriormente y se ha reconocido correctamente.

6. Comparación de versiones

Estas mismas muestras ya habían sido analizadas en 2022, utilizando el sistema de ASR ADITU (Tamayo y Ros-Abaurrea, en prensa). En la siguiente tabla se puede observar la comparación de los datos arrojados por el análisis de ambas versiones:

	ADITU 2022	ADITU 2023
Subtítulos totales	1735	1973
Minutos totales analizados	97	101
Media CPS	16.18	16.5
Media AR con puntuación	94.63 %	94.95 %
Media AR sin puntuación	96.09 %	96.15 %
Errores totales	1967	1975
ER-M	1102	1089
ER-ST	768	802
ER-SE	97	84
Errores por minuto	20.28	19.55
Errores totales sin puntuación	1241	1347
ER-M sin puntuación	431	506
ER-ST sin puntuación	731	771
ER-SE sin puntuación	79	70

Tabla 8: Comparación de ambas versiones

Como puede observarse en la Tabla 8, el número de subtítulos es mayor en el análisis de este trabajo. Este aumento se debe a que en la primera versión no se pudo analizar la muestra 1, debido a que el *software* de entonces no la reconocía. Asimismo, la nueva versión de ADITU ha generado más subtítulos por muestra, aumentando así el total. Por ejemplo, la muestra 2 contaba con 94 subtítulos en el análisis de 2022 y con 106 subtítulos en este análisis de 2023.

Hay que tener en cuenta que, debido a que el número de subtítulos y las muestras totales analizadas no son exactamente los mismos, la comparación de ambas versiones ha de tomarse con cautela. Por lo tanto, que el número de errores sea mayor no significa que la calidad de los subtítulos haya empeorado, puesto que la AR ha aumentado en 0.32 puntos, teniendo en cuenta todo tipo de errores. La media de CPS también ha aumentado en 0.32 puntos y los errores por minuto, por su parte, se han reducido en 0.72 puntos.

6.1. Aspectos que han mejorado

Al comparar las dos versiones, se ha podido observar que en un año el programa ha mejorado en varios ámbitos, y uno de estos aspectos son los números. En el análisis previo (Tamayo y Ros-Abaurrea, en prensa), se describe que el *software* no reconocía el término «*ehuneko*» («por ciento», en castellano), por lo que todos los porcentajes que aparecían en las muestras no se titulaban como tal, creando así errores que confundían al usuario. No obstante, la nueva versión de ADITU reconoce los porcentajes correctamente e incluso utiliza el símbolo adecuado («%») cuando corresponde.

Otra mejoría observada es la de los símbolos. En el análisis previo, cuando el presentador hablaba de la temperatura, la palabra «*graduak*» («grados», en castellano) bien se escribía tal cual o bien el programa la omitía por completo, creando así errores que llevaban a la confusión. En el presente análisis, este problema ya no ocurre.

Por último, ha habido cambios en el programa en cuanto a la puntuación. A pesar de que actualmente el número de errores de puntuación sigue siendo elevado, en el análisis previo se puede observar que es mucho más común el uso de puntos que de comas y que, en la mayoría de los casos, el punto es el signo de puntuación por defecto. Debido a esto, los subtítulos presentaban oraciones más cortas y el uso excesivo de puntos a veces conllevaba errores. Sin embargo, en esta nueva versión, el uso de las comas está más generalizado y el programa detecta con mayor precisión la duración de la pausa del hablante.

6.2. Aspectos que se han mantenido

Aunque ADITU muestre algunas mejoras en su versión de 2023, hay varios problemas que son aún recurrentes. Para empezar, los términos con elementos no propios del euskera siguen sin reconocerse, por lo que no se observa ninguna mejora en el reconocimiento de tildes o fonemas específicos del castellano. No obstante, según Leturia (comunicación personal, en Tamayo y Ros-Abaurrea, en prensa), puede que en un futuro se integren ciertas excepciones a la norma para que el programa pueda reconocer términos que no siguen exactamente las reglas ortotipográficas o fonéticas del euskera, como es el caso de «Athletic».

En relación con este aspecto, el problema con los nombres propios tampoco ha tenido una gran mejora. La mayoría de los nombres propios que no se reconocían en el análisis previo siguen sin reconocerse en el análisis de este trabajo, sobre todo aquellos que cuentan con palabras en otros idiomas o hacen referencia a topónimos extranjeros.

Otro problema aún vigente es el de la identificación de personajes. A pesar de que el uso de las etiquetas ha mejorado en la nueva versión, siguen detectándose errores, y una gran parte de las etiquetas de identificación de personajes que utiliza el programa sigue siendo errónea.

Asimismo, los problemas relacionados con las variantes del euskera siguen ocurriendo. El programa aún sigue reconociendo mejor a los hablantes que utilizan el euskera *batua* y continúa estandarizando la forma de hablar de las personas que se expresan mediante *euskalkis* o que tienen otros acentos que se alejan del *normativo*.

Además, cabe destacar que el programa ha cometido errores en la nueva versión de los subtítulos que no había cometido en la versión anterior. Un ejemplo de esto es el error de continuidad descrito con anterioridad donde se nombraban tres países, pero solo dos de ellos se escribían con mayúscula inicial («*Hungariak, Bulgariak, txekiar errepublikak*», en la muestra 5). El programa había escrito correctamente los tres países en la versión de 2022, pero comete el error en la versión de 2023.

En resumen, podemos concluir que, aunque el *software* de ADITU haya mostrado mejoras en varios ámbitos, sigue sin ser aceptable para su uso en SPS en directo, tomando como referencia la tasa de precisión mínima aceptable del 98 % estipulada por el modelo NER (Romero-Fresco y Martínez Pérez, 2015). Las mejoras observadas están relacionadas con aspectos numéricos y de puntuación, por lo que podemos intuir que el programa ha

evolucionado en aspectos ortotipográficos y que ha sido alimentado con más modelos de lenguaje relacionados con números. Sin embargo, el *software* no ha presentado grandes mejoras semánticas y sintácticas y sigue cometiendo los mismos errores que en el análisis previo. No obstante, debido a la situación minorizada del euskera y los pocos avances en cuanto al reconocimiento del habla que hay en este idioma, los datos obtenidos son prometedores.

7. Conclusiones

El objetivo principal de este trabajo era analizar la calidad de los subtítulos creados a través del sistema de reconocimiento automático del habla ADITU, tomando como base el modelo NER (Romero-Fresco y Martínez Pérez, 2015). Se partía de la hipótesis de que los subtítulos analizados con la versión de 2023 de ADITU serían mejores que aquellos analizados previamente con la versión de 2022 (Tamayo y Ros-Abaurrea, en prensa). Puede concluirse que sí se ha alcanzado el objetivo del trabajo y que se ha rechazado la hipótesis de partida, ya que los datos arrojados por el reconocimiento de ADITU en 2022 y 2023 son muy similares.

Solo una de las 20 muestras ha superado el umbral del 98 % considerado como aceptable según el modelo NER (Romero-Fresco y Martínez Pérez, 2015). Por lo tanto, solo aproximadamente 5 minutos de los 101.67 minutos analizados son considerados aceptables en su totalidad, siguiendo las estipulaciones del análisis.

No obstante, no se deben tener en cuenta únicamente los resultados del análisis cuantitativo. Por ejemplo, una muestra con muchos errores menores (ER-M) es más entendible que una con muchos errores estándar (ER-ST), por mucho que las dos obtengan la misma tasa de precisión según los cálculos del modelo NER. Además, hay otros factores importantes que este modelo no tiene en cuenta, como la velocidad, entre otros, que podrían implementarse en el futuro.

Es importante remarcar las limitaciones de este trabajo, ya que solo se han analizado los subtítulos de 20 muestras (100 minutos en total, aproximadamente) generados por un único programa. Por lo tanto, con estos datos no puede determinarse la calidad del reconocimiento del habla en euskera. Sería recomendable ampliar el minutaje analizado y analizar otros programas de ASR para poder contribuir a los estudios sobre subtitulación en directo a través del reconocimiento del habla.

Asimismo, debemos tener en cuenta las aplicaciones prácticas y las líneas de futuro que abre este TFG. Este análisis, junto con el análisis realizado previamente (Tamayo y Ros-Abaurrea, en prensa), puede resultar de ayuda para que la Fundación Elhuyar observe cuáles son los puntos débiles de este programa de ASR en los subtítulos automáticos en euskera y, con esta información, puedan aplicarse mejoras. Además, teniendo en cuenta que el programa ha implantado ciertas mejoras en un año, es de esperar que en un futuro siga ocurriendo lo mismo.

El mayor aspecto que podría mejorar sería el reconocimiento de otras variantes del euskera, además del euskera *batua*, y el uso de estas a la hora de redactar los subtítulos. Puesto que los programas de ASR trabajan con corpus y diccionarios propios, sería recomendable alimentar el *software* con términos y expresiones propios de los distintos *euskalkis*, para que los subtítulos sean fieles a la realidad lingüística del euskera.

Por último, los datos obtenidos en este trabajo son prometedores para una lengua minoritaria y minorizada como el euskera, así como teniendo en cuenta que los subtítulos han sido generados por un *software* que cuenta con solo tres años de recorrido. Los resultados del análisis se corresponden con otros estudios realizados en otras lenguas oficiales de España, como el gallego (Rico, Docío y García, 2020) y el catalán (Tamayo y de Higes, 2023). No obstante, los datos sobre la cantidad de errores estándar difieren, puesto que se ha observado que estos son más frecuentes en euskera que en inglés (Romero-Fresco y Fresno, 2023), gallego (Rico, Docío y García, 2020) y catalán (Tamayo y de Higes, 2023). Por lo tanto, sería recomendable realizar análisis periódicos de estas mismas muestras para poder observar la evolución del programa ADITU y las mejoras que va implementando.

Referencias

- AENOR (2012). *Norma UNE 153010: Subtitulado para personas sordas y personas con discapacidad auditiva*. AENOR.
- BOE - Agencia Estatal Boletín Oficial del Estado. (8 de julio de 2022). *Ley 13/2022, de 7 de julio, General de la Comunicación Audiovisual*. Ministerio de la Presidencia, Gobierno de España.
- De Linde, Z. y Kay, N. (1999). *The Semiotics of Subtitling*. St. Jerome Publishing.
- Díaz Cintas, J. (2003). *Teoría y práctica de la subtitulación inglés-español*. Ariel.
- Díaz Cintas, J. (2010). «La accesibilidad a los medios de comunicación audiovisual a través del subtitulado y de la audiodescripción». En González, L y Hernández, P. (coord.). *El español, lengua de traducción para la cooperación y el diálogo*, 157-182.
- Díaz Cintas, J. (2012). «Los subtítulos y la subtitulación en la clase de lengua extranjera». *Abehache: Revista da Associação Brasileira de Hispanistas*, Vol. 2, 3, 95-114.
- Díaz Cintas, J. y Remael, A. (2007). *Audiovisual Translation: Subtitling*. St. Jerome Publishing.
- Eugeni, C. (19 de abril de 2023). *Reaching MARS: opportunities for formal and informal training*. [Comunicación en congreso] 8th International Symposium on Live Subtitling and Accessibility, Universitat Autònoma de Barcelona, Barcelona.
- Fresno, N. y Sepielak, K. (2022). «Subtitling speed in Media Accessibility research: some methodological considerations». *Perspectives: Studies in Translation Theory and Practice*, Vol. 30, 3, 415-431.
- Greco, G. M. (2016). «On Accessibility as a human right, with an application to media accessibility». In A. Matamala & P. Orero (Eds.) *Researching audio description. New approaches*, 11-33. Palgrave MacMillan.
- Hlubík, P., Španěl, M., Boháč, M. y Weingartová, L. (2020). «Inserting Punctuation to ASR Output in a Real-Time Production Environment». En Sojka, P., Kopeček, I., Pala, K. y Horák, A. (eds.). *TSD 2020: Text, Speech, and Dialogue*, 418-425.
- Ivarsson, J. y Carroll, M. (1998). *Subtitling*. TransEdit HB.

- Malik, M.; Malik, M. K.; Mehmood, K. y Makhdoom, I. (2021). «Automatic speech recognition: a survey». *Multimedia Tools and Applications*, 80, 9411-9457.
- Martín Sabarís, R., Cantalapiedra, M. J., Zalbidea Bengoa, B. y Aranburu, A. (2016). *Libro de estilo de EITB: información y actualidad*. EITB.
- Pym, A. y Torres-Simón, E. (2021). «Efectos de la automatización en las competencias básicas del traductor: la traducción automática neuronal». En Antoni Vidal Suñe (dir.), Amado Alarcón (dir.). *Ocupaciones y lenguaje: Indicadores y análisis de competencias lingüísticas en el ámbito laboral*, 475-506. Universitat Rovira i Virgili.
- Rico, M., Docío, L. y García, C. (5-6 de noviembre de 2020). *Automatic Galician Subtitles: Towards the Creation of a Live Subtitling Tool*. [Comunicación en congreso] 7th International Symposium on Live Subtitling and Accessibility, Universitat Autònoma de Barcelona, Barcelona.
- Romero-Fresco, P. (2009). «More haste tan speed: Edited versus verbatim respoken subtitles». *Vigo International Journal of Applied Linguistics*, 6, 109-133.
- Romero-Fresco, P. (2011). *Subtitling Through Speech Recognition: Respeaking*. St. Jerome Publishing.
- Romero-Fresco, P. (2016). «Accessing communication: The quality of live subtitles in the UK». *Language & Communication*, Vol. 49, 56-69.
- Romero-Fresco, P. y Fresno, N. (2023). «The accuracy of automatic and human live captions in English». *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 22.
- Romero-Fresco, P. y Martínez Pérez, J. (2015). «Accuracy Rate in Live Subtitling: The NER Model». *Audiovisual Translation in a Global Context*, 28-40. Palgrave Studies in Translating and Interpreting.
- Romero-Fresco, P. y Pöchhacker, F. (2017). «Quality Assessment in Interlingual Live Subtitling: The NTR Model». *Linguistica Antverpiensia*, Vol. 16, 149–167.
- Tamayo, A. y de Higes, I. (19 de abril de 2023). *Automatic speech recognition in Spain: the Basque and Catalan case*. [Comunicación en congreso] 8th International Symposium on Live Subtitling and Accessibility, Universitat Autònoma de Barcelona, Barcelona.

Tamayo, A. y Ros-Abaurrea, A. (en prensa). *Speech-to-text Recognition for the Creation of Live Subtitles in Basque: An Analysis of ADITU Based on the NER Model.*