

# Corpus multimedia *VEIGA* inglés-galego de subtitulación cinematográfica

Patricia Sotelo Dios  
Universidade de Vigo  
psotelod@uvigo.es

## Resumo

Neste artigo presento un proxecto de investigación que consiste na compilación e na explotación do corpus *Veiga*, un corpus multimedia de subtítulos en inglés e en galego. Trátase dun proxecto en fase de desenvolvemento que pretende servir como ferramenta para o estudo e a investigación de certos aspectos relacionados coa práctica da subtitulación intralingüística en inglés e da subtitulación interlingüística do inglés cara ao galego. O *Veiga*, inda que forma parte do corpus paralelo *CLUVI*, transcende o plano textual propio dos demais subcorpus do *CLUVI* e permite observar os subtítulos no seu estado natural, isto é, como parte dun produto audiovisual. Amais de cuestións relacionadas coa construción do corpus e co sistema de buscas, mencionarei algunha das posibles utilidades deste corpus para a práctica, a investigación e a formación en subtitulación.

## 1. *Introdución*<sup>1</sup>

*Non resultaría contradictorio crear unha base de datos ou un corpus de diálogos cinematográficos e os seus correspondentes subtítulos, sen imaxes, coa pretensión de estudar a tradución audiovisual?*<sup>2</sup> (Gambier, 2006).

Non son poucos os proxectos desenvolvidos ata o de agora relacionados coa creación e coa análise de corpus multimedia e multimodais. Ao mesmo tempo, tanto a investigación en tradución audiovisual coma os estudos de tradución baseados en corpus demostran estar gozando de moi boa saúde nesta última década. Con todo, a maioría dos estudos de tradución interlingüística tratan aspectos puramente lingüísticos e culturais, como é a tradución dos distintos rexistros ou do humor verbal, deixando de lado a natureza multisemiótica do texto audiovisual. Concordo con Gambier (2008) en que, se cadra, cumpriría revisar, estender e reformular certos aspectos dos estudos de tradución no tocante á tradución audiovisual, como o concepto de texto e de unidade de tradución e o deseño e a análise de corpus, e non hai dúbida de que precisamos unha nova metodoloxía para tratar fenómenos como a multimodalidade e a multimedialidade. Se consideramos que os filmes, consonte a definición de Kress e van Leeuwen (2001), son articulacións multimodais de múltiples discursos integrados e recoñecemos que non se producen subtítulos para seren lidos de forma

illada, a resposta á pregunta de Gambier é afirmativa: abofé que sería unha contradición crear un corpus de subtítulos formado unicamente por texto cando o que queremos é analizar os anteditos subtítulos tal e como os definen Baker & Saldanha (2009), é dicir, en canto transcricións de diálogos cinematográficos ou televisivos que se presentan de forma simultánea na pantalla, ou estudarmos trazos da subtitulación como son as limitacións espaciais e temporais.

Hogano, inspirados pola crecente abundancia de datos en formato dixital<sup>3</sup> e coa axuda de novos métodos, técnicas e ferramentas informáticos e demais adiantos no deseño de corpus multimedia, tanto investigadores como desenvolvedores de corpus teñen a posibilidade de abordar cuestións e tratar de atopar respostas que resultarían impensables cos métodos máis tradicionais. Porén, para superarmos os obstáculos que inda existen na creación e na explotación de corpus multimedia, cómpre contar cun maior número deste tipo de corpus, nos distintos xéneros, e con máis estudos de carácter tanto teórico como empírico.

Na seguinte sección presentarei os datos dos que consta o *Veiga* neste momento e describerei brevemente os procesos de alíñamento e anotación do corpus e o sistema de busca. O terceiro apartado recolle algunhas reflexións acerca dos posibles usos do devandito corpus en canto ferramenta para a investigación, a práctica e a formación en subtitulación. Menciónanse tamén algunhas das súas limitacións, que serán as que guíen os nosos

<sup>1</sup> Este traballo forma parte do proxecto «Desenvolvemento e explotación de recursos integrados da lingua galega» (INCITE08PXIB302185PR), financiado pola Consellería de Innovación e Industria da Xunta de Galicia.

<sup>2</sup> A tradución das citas é obra da autora do artigo.

<sup>3</sup> Cómpre mencionar a base de datos de subtítulos gratuíta e multilingüe OpenSubtitles.org, a partir da cal se constrúe o corpus OpenSubtitles, que forma parte, pola súa vez, do proxecto OPUS (<http://opus.lingfil.uu.se/index.php>).

pasos máis inmediatos; e rematamos con varias conclusións e apuntamentos relativos aos retos e ao traballo futuros.

## 2. O corpus Veiga

O corpus *Veiga* consta na actualidade de 24 filmes de produción estadounidense, británica e australiana subtítulados en inglés (subtitulación intralingüística<sup>4</sup>) e en galego (subtitulación interlingüística<sup>5</sup>) e distribuídos para DVD, cine e internet. En total contén arredor de 300.000 palabras, inda que está previsto incluír máis filmes, entre eles produtos televisivos, e subtítulos noutras linguas distintas do galego.

Desenvolvido ao abeiro do *CLUVI*<sup>6</sup>, que contén, pola súa vez, varios corpus paralelos de distintos ámbitos (xurídico, científico, xornalístico, tecnolóxico...) e combinacións lingüísticas, o *Veiga* naceu como un corpus textual. Non obstante, en canto contamos coas ferramentas axeitadas para procesar os datos e poñelos a disposición do público, decidimos dar o paso e convertelo nun corpus multimedia.

A diferenza das demais seccións do *CLUVI*, o *Veiga* non se pode considerar estritamente un corpus paralelo, como tampouco se podería catalogar de corpus comparable, segundo a definición tradicional de ambos os conceptos. Baker (1995), por exemplo, emprega o termo «corpus paralelos» para se referir a textos orixinais nunha lingua A e as súas traducións correspondentes nunha lingua B. Por outra banda, Sammut e Webb (2010) definen «corpus comparable» como un conxunto de documentos formado por dous ou máis subconxuntos, cada un nun idioma distinto, mais coa característica común de pertenceren todos a un mesmo campo temático. Un exemplo típico de corpus comparable podería ser un conxunto de artigos xornalísticos redactados en distintas linguas pero que dan conta dos mesmos feitos; non constitúen traducións uns dos outros, mais comparten gran parte de contido semántico.

Xa que logo, o *Veiga* semella ocupar unha especie de «terra de ninguén» entre corpus paralelo e comparable. Dunha banda, os subtítulos en galego non se poden considerar traducións dos subtítulos en inglés, inda que podería ser o caso

que as/os subtituladoras/es partisen da versión inglesa á hora de realizar a tradución ao galego. E, doutra banda, ambas as versións subtítuladas comparten algo máis que contido semántico: as dúas parten dun mesmo texto audiovisual orixinal. Poderíamos dicir, polo tanto, que o orixinal, os subtítulos en inglés e mais os subtítulos en galego manteñen unha relación de tipo triangular. O auténtico e estrito paralelismo sería aquel que se establece entre o texto audiovisual orixinal e cada un dos conxuntos de subtítulos en inglés e en galego. Os subtítulos ingleses corresponderíanse cunha clase particular de transcripción coñecida como subtitulación intralingüística, e os galegos serían froito dunha modalidade de tradución tamén moi peculiar denominada subtitulación interlingüística. Así e todo, non semella desatinado pensar que tamén existe certa relación de paralelismo entre as dúas versións de subtítulos, na medida en que ambas son «subprodutos» do mesmo texto orixinal. En resumo, poderíase establecer un dobre paralelismo unidireccional entre o orixinal e ambos os subtítulos e, asemade, unha correlación bidireccional entre as propias dúas versións de subtítulos, tal e como se mostra na figura 1.

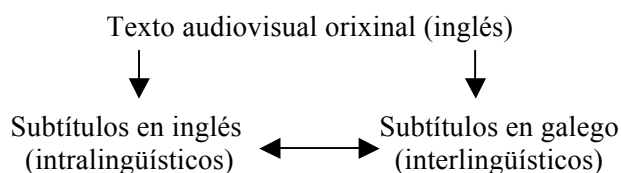


Figura 1: Relación triangular entre orixinal e subtítulos.

Obviamente, cun corpus de subtítulos composto unicamente de texto non sería posible observar paralelismo ningún entre o orixinal e os dous conxuntos de subtítulos. O feito de proporcionármolles ás/aos usuarias/os acceso ao produto audiovisual orixinal permite que estas/es poidan establecer comparacións e paralelismos e, en definitiva, explorar as múltiples dimensións da subtitulación inter- e intralingüística.

### 2.1 Deseño do corpus

Como xa mencionamos, o corpus *Veiga* alóxase no *CLUVI* (Corpus Lingüístico da Universidade de Vigo), que está a ser desenvolvido polo SLI (Seminario de Lingüística Informática). O *CLUVI* funciona como un repositorio de subcorpus paralelos de distintos tamaños e áreas temáticas construídos mediante as mesmas técnicas de procesamento. Esta macroestrutura implica que todas/os as/os desenvolvedoras/es deben seguir os mesmos procesos de aliñamento e de anotación, o

<sup>4</sup> Os subtítulos intralingüísticos están escritos no mesmo idioma que o diálogo orixinal e os seus receptores adoitan ser persoas con problemas auditivos e estudantes de lingua.

<sup>5</sup> Na subtitulación interlingüística, os subtítulos non substitúen o texto orixinal senón que os dous están presentes, de forma sincronizada, no produto subtitulado.

<sup>6</sup> O corpus paralelo *CLUVI* pódese consultar no enderezo <http://sli.uvigo.es/CLUVI/>.

que supón unha vantaxe para elas/es e mais para as/os usuarias/os, que poden así acceder a varios corpus a partir dunha única páxina web, cunha interface e unha páxina de busca e de resultados idénticas.

Non obstante, o *Veiga* multimedia precisa dun procesamento máis complexo: amais de anotarmos fenómenos como as omisións, as adicións e os cambios de orde (pouco frecuentes na subtitulación por mor da necesidade de sincronía) das unidades de tradución<sup>7</sup>, todos os subtítulos inclúen os tempos de entrada e de saída e un indicador de salto de liña, co que as/os usuarias/os poden observar particularidades inherentes á práctica da subtitulación tales como as limitacións espaciais e temporais, a segmentación ou a condensación, entre outras. Alén disto, a versión multimedia do *Veiga*<sup>8</sup> incorpora un capítulo extra: as/os usuarias/os teñen a posibilidade de visualizar os vídeos onde aparecen as unidades de tradución nas dúas linguas, co que accederían ao cotexto na súa forma multisemiótica orixinal. É dicir, que xunta os resultados da súa busca en formato texto aparece unha ligazón aos vídeos cos subtítulos correspondentes en cada unha das dúas linguas (inglés e galego).

Como é de imaxinar, os procesos da creación dos corpus multimedia son bastante complexos e levan moito tempo. Con todo, o imparable avance das tecnoloxías dixitais e da internet abre novos camiños cara ao tratamento de datos multimediais e multimodais. Xa que logo, agardamos poder contar cun maior número de corpus multimedia nun futuro próximo. A seguir, ofrecemos unha breve descrición do procesamento e do sistema de busca do *Veiga* multimedia.

## 2.2 Aliñamento, anotación e edición de vídeo

Bowker (2002) define 'aliñamento' como o proceso de comparar un texto orixinal e a súa tradución, emparellar os correspondentes segmentos e unilos en canto unidades de tradución nunha memoria de tradución. O aliñamento de corpus paralelos fundaméntase, entre outras cousas, na utilización que se lle quere dar ao corpus e no tipo de textos que conforman a base de datos, en función do cal

<sup>7</sup> Véxase Baker e Saldanha (2009) para unha definición rigorosa do concepto de unidade de tradución (UT).

<sup>8</sup> O corpus multimedia de subtítulos *Veiga* está dispoñible no enderezo <http://sli.uvigo.es/CLUVI/vmm.html>. Cómpre lembrar que esta versión multimedia se atopa en proceso de construción e que, neste momento, só se poden consultar 10 dos 24 filmes dos que consta a versión textual.

ha de variar o seu nivel de segmentación (palabras, oracións, parágrafos...).

Como sinalan Guinovart e Sacau (2004), a unidade de segmentación definida para o aliñamento dos textos paralelos do *CLUVI* é a frase ortográfica do texto orixinal. Xa que logo, a correspondencia entre o texto orixinal e a súa tradución sempre será do tipo 1:n. Normalmente, a correspondencia entre oracións do texto orixinal e oracións da tradución é do tipo 1:1. Porén, pódese dar o caso de que a oración orixinal non dispoña de tradución (1:0), ou de que a oración orixinal se corresponda con parte da oración da tradución (1:1/2) ou con dúas oracións na versión traducida (1:2), ou mesmo que unha oración da tradución non se corresponda con ningunha oración no orixinal (0:1). Esta falta de correspondencia entre texto orixinal e traducido (1:0, 0:1) representan omisións ou adicións (sempre respecto do texto orixinal) e etiquétanse mediante unha versión adaptada dos elementos <hi> e <ph> que forman parte da especificación TMX. Con independencia do aplicativo que se empregue para aliñar os textos, todos os subcorpus do *CLUVI* deben seguir a citada especificación. No caso do *Veiga*, empregamos a ferramenta Trans Suite 2000 Align para segmentar e aliñar os textos nas dúas linguas de forma automática, inda que se revisaron todos os aliñamentos e as anotacións foron engadidas de xeito manual.

Nome do filme (UT)	Subtítulo en inglés	Subtítulo en galego
PAR (32)	They said they found Travis.	Atoparon a Travis.
PAR (33)	[[hi type='supr']] Oh, no. [[/hi]]	[[---]]
PAR (34)	What are you going to do?	-E que vas facer?

Táboa 1: Exemplo de omisión.

Nome do filme (UT)	Subtítulo en inglés	Subtítulo en galego
BAB (1201)	In other news...	Noutras noticias...
BAB (1202)	[[---]]	[[hi type='incl']]Aos meus fillos, María Eladia e Eliseo... [[/hi]]
BAB (1203)	[[---]]	[[hi type='incl']]...as luces máis brillantes na noite máis escura. [[/hi]]

Táboa 2: Exemplos de adición.

Á parte disto, debido á natureza particular do noso corpus, tamén se incluíron certos datos

proprios da subtítulos como son a duración e a segmentación dos subtítulos.

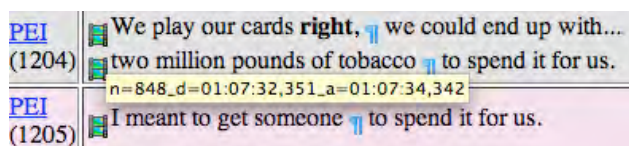


Figura 2: Anotación do código temporal e do salto de liña dos subtítulos.

Na figura 2 móstrase a parte do inglés de dous resultados de busca tal e como se visualizan na web do *Veiga*. A segmentación dos subtítulos vén indicada tanto polo símbolo do caldeirón (¶), que representa un salto de liña dentro dun subtítulo, como pola icona do celuloide, que marca un cambio de subtítulo. Tocante á duración, se pasamos o rato por riba da icona do celuloide poderemos ver os tempos de entrada e de saída do subtítulo en cuestión. A codificación TMX completa correspondente ao aliñamento e á anotación da versión inglesa e galega deste segmento que aparece na dita figura 2 é a seguinte:

```
<tu>
<tuv xml:lang="en">
<seg>[[s n="848" d="01:07:32,351"
a="01:07:34,342"]]We play our cards right,[[l/]]we
could end up with... [[s n="849" d="01:07:34,431"
a="01:07:36,023"]]two million pounds of tobacco[[l/]]to
spend it for us.</seg>
</tuv>
<tuv xml:lang="gl">
<seg>[[s n="808" d="01:07:33,271"
a="01:07:36,946"]]Podemos gastar 2 millóns en
tabaco[[l/]]e pósters de Pamela Anderson. </seg>
</tuv>
</tu>
<tu>
<tuv xml:lang="en">
<seg>[[s n="850" d="01:07:36,511"
a="01:07:39,025"]]I meant to get someone[[l/]]to spend
it for us.</seg>
</tuv>
<tuv xml:lang="gl">
<seg>[[s n="809" d="01:07:37,071"
a="01:07:39,539"]]Buscaremos alguén[[l/]]que o gaste
por nós. </seg>
</tuv>
</tu>
```

Como se pode ver, o código do TMX inclúe unhas indicacións do número (*n*) identificador do subtítulo, dos tempos inicial (*d*) e final (*a*) de cada un dos subtítulos e dos saltos de liña (*l*).

Amais do aliñamento paralelo das unidades de tradución e da anotación, que afecta sobre todo á dimensión textual do corpus, a compilación e o procesamento de material audiovisual require

certos labores de edición de vídeo. Se temos en conta que a) estamos ante un corpus de subtítulos e b) unha das limitacións (e obxecto de estudo) principais da subtítulos é o tempo, un segundo paso lóxico foi o de someter o corpus a un novo proceso de segmentación e aliñamento, en que precisamente as unidades de segmentación estarían constituídas polos propios subtítulos. En primeiro lugar comprobamos que, efectivamente, os tempos dos diferentes subtítulos estivesen ben sincronizados co produto audiovisual correspondente. Nalgúns casos foi preciso editar os subtítulos —para o que empregamos o aplicativo gratuito Subtitle Workshop— e modificar os tempos para axustalos á velocidade de reprodución da película.<sup>9</sup> A seguir, coa ferramenta gratuít VirtualDubMod «pegamos» os subtítulos nas dúas linguas ao produto audiovisual; e, para rematar, editamos cada un destes filmes cos subtítulos nas dúas linguas e segmentámoslos subtítulo a subtítulo. É dicir, todos os textos multimedia do *Veiga* (os filmes orixinais en inglés cos subtítulos en inglés por unha banda e en galego pola outra) foron cortados en clips, cada un dos cales se corresponde cun subtítulo. Como resultado, seguimos a contar con dous conxuntos de subtítulos, en inglés e en galego, compostos cada un deles de tantos clips como subtítulos ten cada produto. Ademais, como moitos dos subtítulos son de pouca duración (un ou dous segundos) e non se poderían visualizar de maneira axeitada, á hora de segmentar os filmes, a cada clip/subtítulo engadíronselle dez segundos extra (os cinco segundos anteriores ao tempo de entrada e os cinco posteriores ao tempo de saída), de xeito que a/o usuario/a poida contar con certo contexto. Logo de obtermos dous conxuntos de videoclips subtítulados para cada un dos textos audiovisuais, estes almacénanse no sistema de ficheiros do servidor identificados polo título da película, a lingua do subtítulado do videoclip (inglés ou galego) e o número do videoclip correspondente<sup>10</sup>, de maneira que cando se realiza unha busca no *Veiga* os resultados mostran non só a unidade de tradución aliñada nas dúas linguas, senón tamén unha ligazón ao vídeo onde aparece ese texto/subtítulo.

En resumo, o corpus multimedia *Veiga* sométese a dous procesos distintos de segmentación: un que afecta só á dimensión textual, isto é, aos subtítulos, e outro que afecta aos

<sup>9</sup> Sobre todo algúns dos subtítulos distribuídos a través da internet, que foron creados para un ficheiro de vídeo cun formato e propiedades concretos que non se corresponden necesariamente co que temos nós.

<sup>10</sup> No caso da figura 2 serían: peixe\_en-848.flv, peixe\_en-849.flv, peixe\_gl-808.flv; e peixe\_en-850.flv, peixe\_gl-809.flv.

subtítulos e mais ao texto audiovisual orixinal ao que estes acompañan. No primeiro caso, a unidade de segmentación é a frase ortográfica (frase en inglés ↔ frase en galego), namentres que no segundo caso a segmentación se realiza ao nivel do subtítulo (texto audiovisual en inglés + subtítulo en inglés ↔ texto audiovisual en inglés + subtítulo en galego), co engadido dos cinco segundos anteriores e posteriores. No primeiro caso a segmentación efectúase de maneira automática, inda que se somete a unha revisión e corrección manual, e no segundo caso, á falta dunha ferramenta que segmente vídeo por subtítulos e engada os anteditos segundos de contexto automaticamente, a segmentación estase a facer de xeito manual. Tocante ao aliñamento, a relación entre as unidades textuais en inglés e en galego é do tipo 1:1, 1:0 ou 0:1, e estas unidades non teñen por que coincidir necesariamente coas unidades de segmentación que constitúen os subtítulos. De feito, o máis habitual é que unha única unidade textual se corresponda con máis dun clip/subtítulo, da mesma maneira que, como mostra a figura 3, un subtítulo en inglés non sempre se corresponde cun único subtítulo en galego.

EAR (340)	Carbon monoxide poisoning, Strychnine, suffocation, breaking the neck, and anal electrocution are some of the more common methods used.	Envenenamento por monóxido de carbono, estricnina, abafamento, quebrar o pescozo, e a electrocución anal son algúns dos métodos usados.
EAR (341)	Removed from his or her cage with a heavy neck-pole, the animal is walked past the rows of bodies of slaughtered foxes, sables, raccoons and wolves, among others.	Sacado da súa gaiola cunha vara no pescozo, os animais camiñan pasando entre ringleiras de corpos de raposos, martas, mapaches, lobos e outros animais mortos.
EAR (342)	Death by anal electrocution is a crude process that requires a probe to be inserted in the rectum while the animal bites down on a metal conductor.	A morte por electrocución anal é un proceso cruel, no cal se insire un electrodo no recto, mentres o animal trava nun condutor de metal.

Figura 3: Exemplo de resultados de busca. A correspondencia entre subtítulos non é do tipo 1:1.

### 2.3 Sistema de busca

Como xa mencionamos con anterioridade, tanto o corpus paralelo *CLUVI* como o *Veiga* se poden consultar en liña a través dun aplicativo PHP deseñado polo SLI. Un dos puntos fortes desta ferramenta de busca reside nas súas múltiples posibilidades de consulta. Permite facer buscas moi complexas de palabras illadas ou grupos de palabras, e mostra as equivalencias bilingües dos termos buscados en contexto. As buscas poden ir en calquera dirección (inglés>galego e galego>inglés) ou nas dúas ao mesmo tempo; é dicir, pódese buscar un termo en cada unha das dúas linguas de forma simultánea. Todos os corpus do *CLUVI* comparten a mesma interface de busca<sup>11</sup>: unha simple caixa onde as/os usuarias/os poden introducir a súa expresión de busca nunha ou en todas as linguas. Por cuestións de dereitos de

autor, o sistema de busca detense logo de atopar as primeiras 1.500 concordancias. De conformidade co dereito de cita, os usuarios en ningún caso teñen acceso aos textos completos, senón que visualizan unicamente aquelas unidades de tradución e os subtítulos correspondentes que coinciden cos seus criterios de busca. Ademais, o texto que se ofrece é o resultado dun procesamento levado a cabo por nós; isto é, non se trata simplemente de texto orixinal, senón que estamos ante material manipulado en maior ou menor grao.

Imos ilustrar este sistema de busca cun exemplo: buscamos pares de unidades de tradución onde apareza tanto a palabra 'for' no inglés como a palabra 'por' no galego. O sistema atopa e mostra unha lista de 131 unidades que coinciden coa busca. Como vemos na figura 4, os resultados aparecen nunha táboa de cinco columnas que conteñen a seguinte información: código do filme e número de unidade de tradución, unidade en inglés, unidade en galego, icona dunha bobina e icona dunha frecha. Se prememos a icona da frecha, ábrese unha fiestra onde aparece o par en cuestión e mais o par anterior e posterior (cotexto). E a icona da bobina enlaza cunha páxina onde podemos visualizar os clips correspondentes tamén de forma paralela, é dicir, nas dúas linguas, e os clips anterior e posterior.







1-CHU (195)	When you're 18, you can go to hell for all I care.	Canda teñas 18, por mín como se vas para o inferno,		
2-CHU (258)	I've been a prisoner of my love for you for a very long time.	Levo moito tempo presa no amor que sinto por ti.		
3-CHU (436)	-I'd do anything for you.	- Faría calquera cousa por ti,		

Figura 4: Exemplo de resultados de busca.

### 3. Posibles usos e limitacións do corpus multimedia Veiga

Xa o dicía Baldry (2004): precisamos acceder aos textos na súa forma *in vivo*, de xeito que a relación entre a pista de son e de vídeo se mantenga intacta, porque a maneira principal en que un texto filmico constrúe o seu significado é mediante a sincronización entre as fontes sonora e visual. Con todo, antes de falarmos das posibles vantaxes do *Veiga*, comecemos por recoñecer cales son as súas limitacións máis evidentes.

#### 3.1 Limitacións

O primeiro hándicap atopámolo no reducido tamaño do corpus. No noso favor podemos alegar que, neste momento, só hai dúas persoas traballando no proxecto e que está previsto ampliálo con filmes subtitulados emitidos pola televisión e poida que se inclúan outras linguas. Non cabe dúbida de que canto maior sexa o corpus, máis variado e extenso é o conxunto de fenómenos

<sup>11</sup> Cómpre indicar que o *Veiga* multimedia está en proceso de emigrar a un novo sitio web.

que podemos atopar, o que afectaría positivamente a fiabilidade dos posibles estudos baseados nos datos do *Veiga*. Non obstante, tal e como afirman McEnery & Wilson (2001), o tamaño non ten por que ser garantía de representatividade. Ademais, un corpus pequeno pero especializado tamén pode ser útil na investigación de fenómenos particulares.

Unha segunda limitación podería ser a heteroxénea procedencia e autoría dos subtítulos traducidos, o que semella demandar un subsecuente cambio de enfoque á hora de observar os datos e suscita a cuestión da calidade da tradución (e do corpus). Como xa se comentou, os ficheiros dos subtítulos en galego foron distribuídos en DVD, no cinema e na internet. En concreto, sete deles foron creados para DVD, é dicir, que é bastante probable que os seus autores fosen tradutoras/es profesionais e que pasasen algún tipo de control de calidade. Catorce deles foron creados para a gran pantalla e proxectados en ciclos de cine. Neste caso, a maior parte das/os subtítuladoras/es son voluntarias/os<sup>12</sup>, e estarían a medio camiño entre a primeira caste de tradutoras/es (profesionais) e a que vén a seguir (afeccionadas/os). E os outros tres grupos de subtítulos constituirían casos dun novo tipo de subtitulación que se vén practicando en España, e mais noutros países, coñecido no mundo anglosaxón co nome de *amateur subtitling* e descrito como unha práctica desenvolvida por non profesionais e que se rexe por unhas limitacións completamente diferentes ás da subtitulación profesional. O resultado final depende en gran medida do coñecemento desta/e afeccionada/o da lingua orixinal, o que probablemente derive en abundantes erros e interpretacións equivocadas. Así e todo, a calidade non foi un criterio que tivésemos en conta cando recompilamos o corpus.

E a terceira eiva constitúena os xa citados labores de procesamento e de edición do material audiovisual deste corpus, que, malia todos os adiantos tecnolóxicos dos que dispoñemos na actualidade, seguen a provocar unha gran demora no avance do proxecto.

En consecuencia, con independencia do uso que se queira facer dos resultados das buscas no corpus *Veiga*, cómpre ter sempre presentes as anteditas limitacións.

### 3.2 Posibles usos

Malia todo o anterior, pódense mencionar varias posibles utilidades do *Veiga* multimedia. Por unha

banda, pódese empregar como un banco de exemplos, unha base de datos que as/os investigadoras/es poden usar para ilustrar os seus traballos e probar as súas hipóteses.

No eido pedagóxico, pódese usar en varios contextos, dende en cursos de lingua que versen sobre os rexistros e a xerga ata en cursos especializados en tradución audiovisual e subtitulación, pasando por cursos de tradución non necesariamente especializada onde se trate o transvasamento de referentes culturais, frases feitas, coloquialismos, etcétera (Zanettin *et al.*, 2003). Na nosa opinión, é fundamental que as/os docentes de tradución audiovisual lle ofrezan ao seu alumnado material auténtico que se preste a análises contrastivas de textos orixinais e textos traducidos. Por outra banda, o uso de corpora como metodoloxía do ensino da tradución favorece o desenvolvemento de espírito crítico, aprendizaxe autónoma, toma de decisións e outras competencias tanto propias dos estudos de tradución como transversais (Beeby *et al.*, 2009). Ao mesmo tempo, tamén pode servir de ferramenta para a aprendizaxe en liña, xa que as/os alumnas/os contarían coa posibilidade de explorar as propiedades textuais en tanto que oen e ven os vídeos, que, amais, poden reproducir e pausar cando o desexen.

Para rematar, tamén os profesionais da subtitulación poderían facer uso deste corpus en canto unha colección de subtítulos xa feitos, no sentido de que poderían ver as solucións polas que optaron outras/os profesionais para resolver certas limitacións ou dificultades particulares da subtitulación.

Como xa comentamos, o tamaño limitado do corpus e a natureza híbrida dos subtítulos traducidos non permiten establecer xeneralizacións sobre a práctica da subtitulación inter- e intralingüística. De feito, cabería mesmo trazar máis distincións en función do xénero dos textos audiovisuais (películas, documentais, filmes para as/os cativas/os...) e do medio de distribución do produto. Emporiso, cómpre aclarar que a nosa pretensión é só fornecer unha ferramenta que poidan usar tanto investigadores como profesionais e docentes para ilustrar aspectos particulares da subtitulación. Por unha banda, o *Veiga* permite observar cuestións técnicas como a presentación dos subtítulos na pantalla (número de liñas, posición, cor, marcas de diálogo) e a súa duración (tempos de entrada e de saída, intervalo<sup>13</sup>, cambios de plano e sincronización). Segundo Díaz Cintas e

<sup>12</sup> Son cinéfilas/os e afeccionadas/os ás versións orixinais subtituladas e forman parte da propia asociación organizadora dos ciclos de cine, mais non recibiron formación especializada en tradución para o subtítulo e realizan este labor de balde.

<sup>13</sup> De Linde e Kay (1999) distinguen entre *leading* e *lagging*. Cando o subtítulo precede o texto oral falamos de anticipación e cando o segue faláramos de demora.

Remael (2007), a práctica da subtitulación é bastante heteroxénea e varía de maneira considerable dun programa audiovisual, dunha empresa e mesmo dun país a outro. Malia os varios intentos que se fixeron de elaborar un conxunto de convencións ou pautas harmonizadas, semella que en cada idioma/cultura se subtitula consonte a práctica tradicional que lle é propia. Por outra banda, os subtítulos, tanto os intralingüísticos coma os interlingüísticos, constitúen versións condensadas do texto audiovisual. A subtitulación adoita implicar unha selección do material lingüístico, o que obriga ás/aos subtituladoras/es a tomaren continuas decisións acerca do que resulta importante e do que semella superfluo ou mesmo redundante. A redundancia, xa que logo, é un concepto esencial na subtitulación, xa que parte da información omitida nos subtítulos pode ser proporcionada por outros elementos do texto audiovisual, como a imaxe e/ou o son. Con todo, a redución é un fenómeno frecuente e decote inevitable que implica omisión de información. O *Veiga* multimedia non só sitúa os subtítulos ao carón do texto audiovisual orixinal, senón que coloca cara a cara as dúas versións subtituladas, co que podemos observar fenómenos, tales como a cohesión e a condensación, enraizados na semiótica da subtitulación.

#### 4. Conclusións e traballo futuro

Nestas páxinas presentamos o corpus multimedia de subtítulos inglés-galego *Veiga*, un proxecto en fase de elaboración que tenta facerse eco dunha idea formulada por varias/os autoras/es nos últimos anos: cómpre superar o enfoque tradicional da elaboración de corpus, onde prima o texto escrito, e convencerse da necesidade de deseñar corpora multimedia que amosen os aspectos polisemióticos do discurso filmico e da subtitulación. Tamén tratamos cuestións relacionadas cos datos e coa construción do corpus; en concreto, coa condición «intermedia» dos subtítulos, que non son exactamente traducións uns dos outros. Comentamos algunha das eivas máis obvias do corpus como son o seu tamaño e as limitacións tecnolóxicas, que agardamos poder resolver no futuro. E, para rematar, mencionamos varias áreas da práctica, da investigación e da docencia da subtitulación onde o *Veiga* podería ser de certa utilidade.

O primeiro chanzo no camiño consistirá en incluír no sitio web datos acerca do medio de distribución (DVD, cinema, internet...) e/ou autoría dos distintos subtítulos, de maneira que as/os usuarias/os poidan facer un uso o máis informado posible do corpus. E como pasos futuros, amais de completar a transmutación dos subtítulos en

formato texto ao formato audiovisual, e de aumentar o número de filmes subtitulados para o cinema e daqueles creados por afeccionados e distribuídos a través da internet, temos a intención de incluír tamén produtos televisados —emitidos principalmente pola CRTVG—. Deste xeito estaremos ampliando o tamaño do corpus, engadindo un novo modo de distribución —que se sumaría ao DVD, ao cinema e á internet— e diversificando o seu contido, xa que é moi probable que algúns deses novos produtos constitúan novos xéneros dentro dos textos audiovisuais. Tocante á extensión a outras linguas, considerámolo como unha posibilidade a máis longo prazo que habería que estudar chegado o momento; polo de agora, a nosa prioridade é seguir a traballar na combinación inglés-galego e mellorar o corpus de maneira que cada día sexan menos as súas limitacións e máis as súas aplicacións e posibilidades.

#### Referencias

- Baker, M. 1995. Corpora in Translation Studies: An Overview and Some Suggestions for Future Research. *Target*, 7(2):223-243.
- Baker, M. e G. Saldanha, editoras. 2009. *Routledge Encyclopedia of Translation Studies*, segunda edición, páxinas 244-245 e 304-306. Londres e Nova York: Routledge.
- Baldry A. e C. Taylor. 2004. Multimodal concordancing and subtitles with MCA. En A. Partington, J. Morley e L. Haarman, editores, *Corpora and Discourse*, páxinas 57-70. Bern: Peter Lang.
- Beeby, A., P. Rodríguez Inés e P. Sánchez-Gijón, editoras. 2009. *Corpus Use and Translating: Corpus Use for Learning to Translate and Learning Corpus Use to Translate*. Amsterdam-Philadelphia: John Benjamins Publishing.
- Bowker, L. 2002. *Computer-Aided translation Technology: A Practical Introduction*. Ottawa: University of Ottawa Press.
- De Linde, Z. e N. Kay. 1999. *The Semiotics of Subtitling*. Manchester: St. Jerome Publishing.
- Díaz Cintas, J. e G. Anderman, editores. 2009. *Audiovisual translation: Language Transfer on Screen*. Londres: Palgrave Macmillan.
- Díaz Cintas, J. e A. Remael. 2007. *Audiovisual Translation: Subtitling*. Manchester: St. Jerome Publishing.
- Gambier, Y. 2008. Recent developments and challenges in audiovisual translation research. En D. Chiaro, C. Heiss e C. Bucaria, editoras, *Between Text and Image: Updating Research in Screen Translation*, páxinas 11-33. Amsterdam-Philadelphia: John Benjamins Publishing.
- Gambier, Y. 2006. Multimodality and Audiovisual Translation. En *MuTra 2006: Audiovisual*

- Translation Scenarios: Conference Proceedings*. Copenhagen. Disponible en liña no enderezo [http://www.euroconferences.info/proceedings/2006\\_Proceedings/2006\\_Gambier\\_Yves.pdf](http://www.euroconferences.info/proceedings/2006_Proceedings/2006_Gambier_Yves.pdf).
- Gómez Guinovart, X. e E. Sacau Fontenla. 2004. Parallel corpora for the Galician language: building and processing of the CLUVI (Linguistic Corpus of the University of Vigo). En T. Lino *et al.*, editoras, *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC 2004)*, páxinas 1179-1182. Lisboa. Disponible en liña no enderezo <http://webs.uvigo.es/sli/arquivos/lrec2004.pdf>.
- Kress, G. e T. van Leeuwen. 2001. *Multimodal Discourse: The Modes and Media of Contemporary Communication*. Nova York: Oxford University Press.
- McEnery, T. e A. Wilson. 2001. *Corpus Linguistics*, segunda edición. Edimburgo: Edinburgh University Press.
- Sammut, C. e G. I. Webb, editores. 2010. *Encyclopedia of Machine Learning*. Nova York: Springer.
- Zanettin, F., S. Bernardini e D. Stewart, editores. 2003. *Corpora in Translator Education*. Manchester: St Jerome Publishing.