APERTIUM Y LA TRADUCCIÓN AUTOMÁTICA BASADA EN REGLAS

CREACIÓN DE UN DICCIONARIO FANÉS PARA LA PAREJA DE IDIOMAS ITALIANO-FANÉS.

Federico Gambini

Trabajo de final de máster

Director: Adrià Martín-Mor

Facultat de Traducció i Interpretació, Màster en Tradumàtica, 2019

Datos del TFM

Título (es): *Apertium y la traducción automática basada en reglas. Creación de un diccionario fanés para la pareja de idiomas italiano-fanés.*

Títol (ca): *Apertium i la traducció automàtica basada en regles. Creació d'un diccionari fanés per a la parella d'idiomes italiano-fanés.*

Title (en): *Rule-based machine translation with Apertium. Creation of a Fanés dictionary for the Italian-Fanés language pair.*

Autor: Federico Gambini

Tutor: Adrià Martín-Mor

Centro: Facultad de Traducción e Interpretación

Estudios: Máster en Tradumàtica: Tecnologías de la Traducción

Curso académico: 2018-2019

Palabras clave (es)

Apertium, traducción automática basada en reglas, TABR, lenguas minoritarias, fanés, digitalización, código libre.

Resumen (es)

En este trabajo de final de máster he puesto las bases para la creación de un traductor automático basado en reglas a través del software de código libre Apertium, desarrollado por la Universidad de Alicante. La pareja de idiomas en cuestión es italiano-fanés. En el marco teórico he hablado del fanés, un idioma minoritario y variedad lingüística del romañol, que se habla en la parte norteña de la región italiana de las Marcas. Luego, he definido la traducción automática y más en específico la basada en reglas y sus características. Finalmente, he desarrollado la preparación del entorno Ubuntu en Windows para instalar Apertium y he ilustrado mi flujo de trabajo en la creación de las entradas en los diccionarios que compondrán el traductor automático basado en reglas. Teniendo que crear un diccionario lingüístico digital fanés desde cero, me he enfrentado a muchas dificultades ya que no tenía conocimientos previos de Apertium y además porque el fanés no tiene una gramática oficial y una academia lingüística reguladora. No obstante, he podido comprobar que casi todas las entradas que he insertado funcionan y que quizás en un futuro no muy lejano, podrán incrementar el número de entradas para finalmente ser completado.

Paraules clau (ca)

Apertium, traducció automàtica basada en regles, TABR, llengües minoritàries, fanés, digitalització, codi lliure.

Resum (ca)

En aquest treball de final de màster he posat les bases per a la creació d'un traductor automàtic basat en regles a través del programari a codi lliure Apertium en Windows, desenvolupat per la Universitat d'Alacant. La parella d'idiomes en qüestió és italià-fanés. En el marc teòric he parlat del fanés, un idioma minoritari i varietat lingüística del romanyès, que es parla en la part del nord de la regió italiana de les Marques. Després,

he definit la traducció automàtica i més en específic la basada en regles i les seves característiques. Finalment, he desenvolupat la preparació de l'entorn Ubuntu per a instal·lar Apertium i he il·lustrat el meu flux de treball en la creació d'entrades en els diccionaris que compondran el traductor automàtic basat en regles. Havent de crear un diccionari lingüístic digital fanés des de zero, m'he enfrontat a moltes dificultats ja que no tenia coneixements previs d'Apertium i a més perquè el fanés no té una gramàtica oficial ni una acadèmia lingüística reguladora. No obstant això, he pogut comprovar que gairebé totes les entrades que he inserit funcionen i que potser en un futur pròxim es podran incrementar el número d'entrades per finalment completar-lo.

Keywords (en)

Apertium, rule-based machine translation, RBMT, minority languages, Fanés, digitalization, open source.

Abstract (en)

In this master's final project I have laid the foundations for the creation of a rule-based machine translation through the open source software Apertium, developed by the University of Alicante. The language pair I worked on is Italian-Fanés. First, I talked about Fanés, a minority language and linguistic variety of Romagnol, spoken in the northern part of the Italian region of the Marche. Then, I have defined what is machine translation and more specifically, the one based on rules and its characteristics. As a last part, I described the preparation of the Ubuntu environment on Windows to install Apertium and I illustrated my workflow in the creation of the entries in the dictionaries that will build up the machine translator. Having to create the first digital dictionary of Fanés from scratch, I have faced many difficulties since I had no prior knowledge of Apertium and also because Fanés does not have an official grammar and a regulatory linguistic academy. However, I have been able to verify that almost all the entries that I have been able to insert they do work, and that perhaps in the future the number of entries may be increased and finally completed.

Índice de contenido

Índice de ilustraciones y tablas

1. **Introducción**

El fanés es una lengua que se habla en la provincia de Pesaro y Urbino, más en específico en la ciudad de Fano, que consta en fecha 31 de diciembre de 2017, de 60 978 habitantes. No tiene muchos recursos lingüísticos y/o escritos y necesita una estandarización. En la asignatura de Traducción de Productos Digitales del Máster de Tradumática: Tecnologías de la Traducción, llevé a cabo una localización íntegra de la aplicación de mensajería de Telegram para Android al idioma fanés. Esto ha resultado también en una creación de una primera memoria de traducción del par de idiomas italiano-fanés. En octubre de 2018 se presentó en Fano un proyecto de una Wiki comunitaria, donde cada usuario que se haya inscrito, puede aportar sus conocimientos y contribuir a la creación de una enciclopedia en línea. En mi caso, he decidido pues de seguir adelante con este intento de digitalización y estandarización de mi segunda lengua madre con el trabajo de final de máster. He elegido entonces de intentar crear un motor de traducción automática italiano-fanés a través de Apertium, una plataforma de código libre. Tendré que crear desde cero un diccionario monolingüe fanés y un bilingüe (italiano-fanés).

Este trabajo es una ocasión para enriquecer mis conocimientos informáticos, de traductor y de lingüista y al mismo tiempo, para empujar el proceso de salvaguardia de los idiomas minoritarios de Italia.

1.1 Objetivos

Actualmente para el idioma fanés, los recursos lingüísticos (documentos escritos) como los tecnológicos (corpus, memorias de traducción, bases de datos) son escasos. Al momento, la falta de textos escritos de acuerdo con las reglas ortográficas y léxicas, hace necesario optar por un sistema de traducción automática basado en reglas de transferencia y diccionarios escritos en lenguaje de marcado.

El objetivo de este trabajo es intentar crear las bases para construir un motor de traducción automática, de italiano-fanés, a través de la plataforma a código libre Apertium. Se trata de un sistema que se adapta bien a la traducción entre pares de lenguas que pertenecen a la misma raíz lingüística (lenguas romances), en mi caso el italiano y el fanés. Este proyecto podrá también sentar las bases para que, en un futuro inmediato, se pueda trabajar en la traducción de otros pares de idiomas como el fanés-catalán y el fanés-español. También podría enriquecer la documentación digital del fanés, que hasta ahora se limita a una comunidad Wiki en la web, una localización integra de Telegram para Android, y de una memoria de traducción sacada de esta última.

Aunque llevar el cabo este proyecto requiere mucho tiempo y muchos conocimientos informáticos, he tomado esta ocasión para aprender a programar en lenguaje XML el cual estoy seguro que me resultará muy útil para mi futuro profesional.

Para el flujo de trabajo, se pretende crear un corpus en italiano, del cual quisiera intentar sacar una lista de frecuencia de palabras. De esta lista podré tener una idea a cuáles palabras daré la antelación para crear las entradas en los diccionarios. Para aprender a utilizar Apertium, intentaré apoyarme a las guías presentes en la web y a los archivos de diccionarios ya existentes de otras parejas de idiomas.

Como línea de futuro, intentaré contactar con el equipo de Apertium para que se pueda tener en consideración esta pareja de idioma y publicarla. Al mismo tiempo me gustaría enseñar este proyecto a los medios de comunicación de Fano, para que quizás, alguien pueda colaborar conmigo para completar la pareja de idioma italiano-fanés.

2. Marco teórico y antecedentes

2.1 Situación de los idiomas de Italia y el fanés.

La dialectología italiana es una disciplina de investigación específica pero también científica. Los primeros estudios nacieron en 1873 con el "*Saggi ladini*" de Graziadio Isaia Ascoli, publicado en su propia revista "*Archivo glotológico italiano*". En las últimas décadas del siglo XIX hasta hoy muchas colecciones sistemáticas de información sobre dialectos como atlas lingüísticos, compilaciones de vocabularios dialectales, recopilación de textos y diversos documentos, hasta un proyecto internacional llamado "Carta dei Dialetti Italiano", (comúnmente abreviado como CDI).

Históricamente, la palabra dialecto viene del griego *diàlektos* que significa "conversación" pero también 'idioma de un pueblo en particular'. Luego se tradujo al latín en las formas *dialectos* o *dialectus* que significa 'discurso local tomado en importancia literaria' (Cortelazzo Manlio, 1969).

Actualmente el término se designa para indicar una variedad lingüística delimitada territorialmente que vive en convivencia con el medio de comunicación dominante, el italiano (Avolio Francesco, 2009).

Ya que no existe una autoridad lingüística que regule la estandarización y que permita la enseñanza en las escuelas de muchas de las lenguas minoritarias de Italia, hoy en día los jóvenes suelen hablar solamente italiano, mientras la mayoría de los ancianos siguen siendo bilingüe. (Marcato Carla, 2007).

Delimitar los idiomas de Italia siempre ha estado difícil durante los años porque como por ejemplo en el mismo sistema sardo podemos encontrar variedades dialectales, en cada una de las otras se encontrarán muchas otras (Graffi y Scalise, 2003).

Para intentar recoger rasgos lingüísticos en común (que estos sean fonéticos, morfosintácticos o léxicos) se utilizan las isoglosas. Tal y como define el Portal de Lingüística Hispánica, se trata de una "línea imaginaria con que se divide un territorio de manera geográfica según el uso de un rasgo lingüístico concreto. No es absoluta, dado que la lengua es un elemento en constante evolución"[1]. En Italia, las isoglosas más

[1] http://hispaniclinguistics.com/glosario/isoglosa/

importantes son las de Ancona-Roma y las que cruzan los Apeninos y se mueven de La Spezia-Massa Carrara al área entre Rimini-Fano. Estas última juntan la zona lingüística del norte, dejando fuera la Toscana y los dialectos centro-meridionales (Balducci, 1984).

En referencia al idioma de Fano, se trata de una variedad lingüística del romañol, una lengua galorromance que pertenece al grupo emiliano-romañol. Aunque Fano (y su provincia de Pesaro y Urbino) pertenece políticamente a la región de Las Marcas, los rasgos lingüísticos del idioma son más similares a la región de la Romaña (Balducci, 1984).

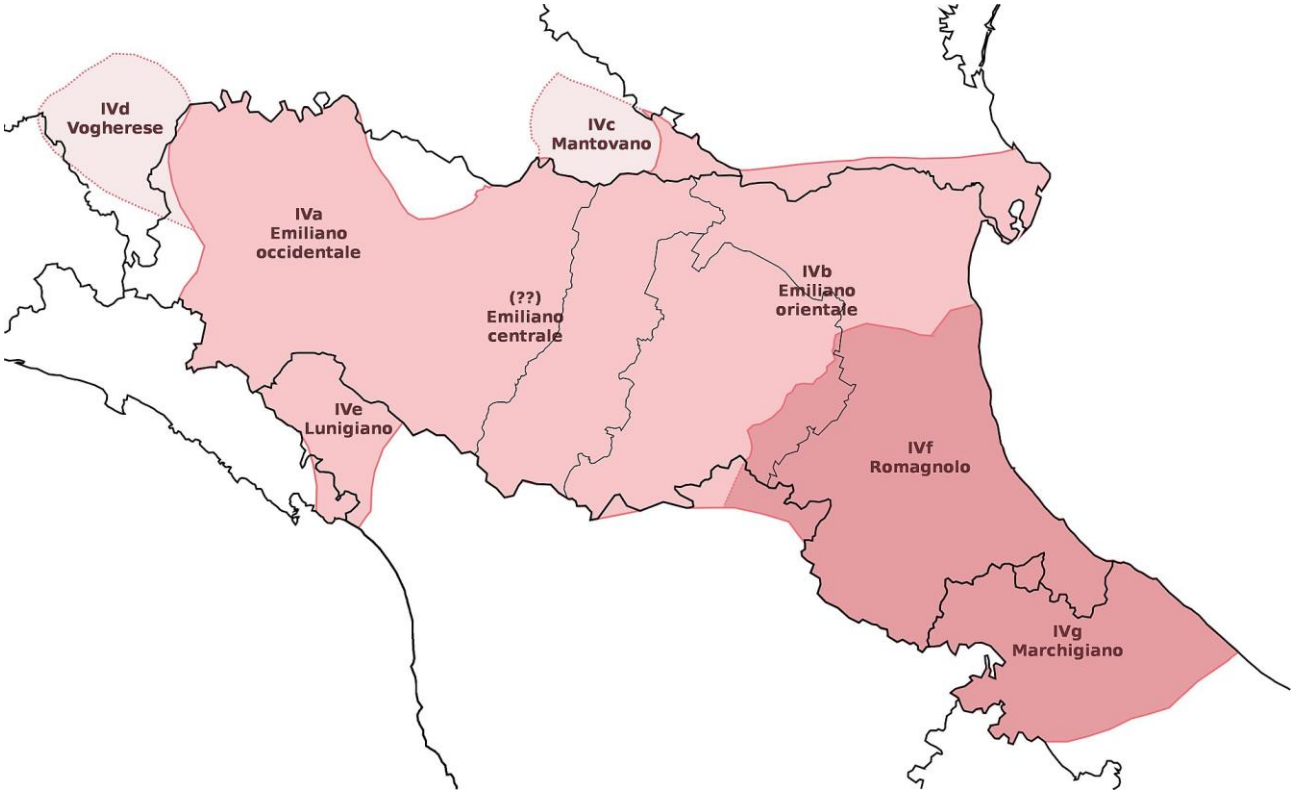

Ilustración 1. Variedades del Emiliano-romañol

La cultura de Fano cuenta con numerosos poetas, novelistas, compañías teatrales, incluso directores de cine, que Intentan dar valor a la cultura y al idioma de Fano. Dos obras importantes que han ayudado a motivar a otros académicos en compilar un diccionario lingüístico son: *Le parol de Fan: raccolta di vocaboli* y *locuzioni caratteristiche del dialetto fanese*, publicadas en 1975 por Sperandini y Vampa. Representan el primer intento de una construcción de un glosario de Fano, que puede contar con cien páginas de entradas.

El primer verdadero diccionario fanese se publicó en la ciudad del mismo nombre en 1992: *Come Parlano i Fanesi: Vol. I 'Dizionario', Edizione la Fortuna*. Los autores,

Agostino Silvi y Ermanno Simoncelli, han seguido con la investigación y los estudios sobre el dialecto de su ciudad con éxito: en 2004 se publicó una segunda edición, más completa y actualizada, que incluye una versión italiano-fanese y un apéndice gramatical.

2.1 La traducción automática basada en reglas

En 1948 el investigador e ingeniero Warren Weaver (Hutchins y Somers, 1992) propuso crear un programa capaz de traducir un texto de un idioma a otro sin la intervención del hombre. En el documento titulado “Traducción”, escrito para la División de Ciencias Naturales de la Fundación Rockefeller, Warren formuló algunas hipótesis sobre los potenciales y métodos de TA: apoyó la validez del método de reemplazo palabra por palabra y propuso integrarlo con técnicas estadísticas. Se aplica para detectar la frecuencia de palabras y caracteres en textos paralelos. La idea de Weaver logró captar la atención de varias empresas en muy poco tiempo, lo que decidió financiar el proyecto.

En específico la traducción automática (TA) se trata con textos escritos o bien, informatizados (Forcada Mikel L., 2009). Podemos hablar entonces de una transformación, utilizando un sistema informático, de un texto escrito en la *lengua de origen*, a otro texto escrito en la *lengua meta.*

Las traducciones resultantes de este sistema de transformación suelen ser menos precisas que las hecha por profesionales, ya que hay que tener en cuenta entre otras cosas la *ambigüedad* de los textos producidos por humanos y otros problemas descritos y divididos por Arnold en los siguientes grupos (Arnold, D., 2003 en Carme Armentano-Oller, Antonio M et al., 2007:3):

- *La forma no determina completamente el contenido.* En este caso se habla de *ambigüedad:* un humano puede llegar a entender el sentido de un texto a través del contexto mientras es difícil hacer que un programa llegue a hacerlo. Lo que destacan Carme Armentano-Oller, Antonio M et al., es que los humanos tienen *conocimiento del mundo* y que es complicado sistematizarlo en un programa de ordenador.

- *El contenido no determina completamente la forma.* Ya que hay muchas maneras de expresar una misma cosa en un idioma, se deben de crear y aplicar estrategias que reduzcan las varias formas de decir lo mismo, para que un ordenador no tenga que enfrentarse a estos tipos de complejidades.

- *Distintas lenguas usan estructuras diferentes para expresar las mismas cosas.* En este caso se puede decir que hay idiomas como el inglés que no utilizan artículos en una frase como “I like videogames” donde en castellano “me gustan los

videojuegos”. Esto para decir que entre idiomas hay estructuras muy diferentes que complican la traducción directa de un motor de traducción automática.

Se pueden distinguir dos usos de la TA. El primero es la *asimilación.* En este caso, la TA sirve como medio para obtener una idea general del texto de origen. Su uso es inmediato y superficial, ya que luego las traducciones no se conservarán. El sentido del texto tiene más importancia que los errores que hay en la traducción (Forcada, 2009).

Mientras tanto, el uso más importante de la traducción automática es, como denomina Mikel Forcada, la *disseminació:*

> *«Es diuen així perquè comporten* l'ús de la traducció automàtica com a pas intermedi en la producció d'un document en la llengua meta que serà publicat o disseminat; per tant, la traducció en brut es conserva perquè l'ha de revisar i corregir, o com se sol dir, posteditar, una persona especialitzada. Simplificant, podem dir que la traducció automàtica seguida de postedició constituirà una alternativa a la traducció professional només si el seu cost conjunt és menor que el de la traducción professional tradicional.» (Forcada, 2009:16).

Existen varios tipos de tecnología de traducción automática, y actualmente se pueden clasificar en dos grandes grupos: traductores automáticos basados en reglas y traductores automáticos basados en corpus.

La traducción automática basada en corpus es la que utiliza un gran número de textos y/o frases bilingües alineándolos con la traducción correspondiente en el otro idioma. Actualmente es la que se suele utilizar más y dentro de este grupo se encuentran los sistemas basados en ejemplos, los estadísticos y los neuronales (Ginestí-Rosell y Forcada, 2009).

De otra parte, la traducción basada en reglas (TABR) es un sistema que se basa en las informaciones lingüísticas de la lengua de origen y la lengua de llegada que se sacan de diccionarios monolingües, bilingües o multilingües informatizados. Un rol importante lo lleva la gramática, que tendrá que cubrir las reglas principales de los dos idiomas. De hecho, el sistema de TABR genera las oraciones después de pasar a través de procesos de comprobación léxica, morfológica y sintáctica, creados por un humano. Sin embargo, requiere un gran esfuerzo de desarrollo, pero funciona bien entre lenguas cercanas y con pocos recursos. Se suelen distinguir tres principales componentes: un *motor* (que sirve para descodificar y recombinar), *datos* (datos lingüísticos o corpus paralelos) y

herramientas para mantener los datos y convertirlos en un formato que pueda leer la máquina (Forcada, 2009).

2.2 El traductor automático de la plataforma Apertium

La *Free Software Foundation*[2] es una organización sin fines de lucro, con el propósito de difundir la promoción del software libre. Ella misma presenta los criterios que califican si un software se puede considerar de código libre o no[3]. Para ser clasificado como software libre, un programa tiene que respetar las libertades de la comunidad y de los usuarios. Más en específico, los usuarios deben de tener la oportunidad de ejecutar, copiar, distribuir, cambiar y mejorar el software. El todo se resume en cuatro libertades esenciales:

- La libertad de ejecutar el software cuando quieran, con cualquier propósito (libertad 0).
- La libertad de estudiar cómo funciona el software, y modificarlo para que funcione cómo quieran (libertad 1). El acceso al código fuente es un prerrequisito para ello.
- La libertad de redistribuir copias para ayudar a los demás (libertad 2).
- La libertad de distribuir copias de sus versiones modificadas a los demás (libertad 3). Haciendo esto pueden dar a toda la comunidad la oportunidad de beneficiarse de sus cambios. El acceso al código fuente es un prerrequisito para ello.

La razón por la cual los números van de 0 a 3 es histórica. En los años 90, había 3 libertades, la 1, 2 y 3. Luego se tomó en consideración que la libertad de ejecutar el programa necesitaba una mención explícita. Ya que era más esencial que las otras tres, debía de estar en una posición precedente. Entonces, en lugar de enumerar las otras de nuevo, se decidió ponerla como número 0 (Free Software Foundation).

Apertium es un sistema de traducción automática de código libre creado por la Universidad de Alicante en 2004. Se basa en la filosofía Unix, es decir que en su interior hay diferentes programas (denominados módulos) que funcionan individualmente pero que en conjunto completan la tarea de traducción. Para guardar las informaciones lingüísticas utiliza el formato XML. Inicialmente estaba concebido sólo para parejas de idiomas románicos, pero durante los años se ha ido expandiendo su uso a idiomas de raíces diferentes (inglés-catalán). El diseño oficial está basado en los sistemas que

[2] https://es.wikipedia.org/wiki/Free_Software_Foundation

[3] https://www.gnu.org/philosophy/free-sw.en.html

habían desarrollado el grupo Transducens de la Universitat de Alicante, interNOSTRUM[4] (Forcada Mikel L., 2009).

Para generar traducciones que sean razonablemente inteligibles y fáciles de corregir entre lenguas relacionadas como el español y el catalán o el portugués, solo hay que mejorar la traducción palabra por palabra con: procesamiento léxico robusto (incluyendo unidades léxicas multi-palabra), desambiguación léxica categorial (*parte-of-speech tagging*) y procesamiento estructural local basado en reglas simples y muy formuladas para transformaciones estructurales frecuentes (Forcada Mikel L., 2009).

El programa para los desarrolladores es ejecutable sólo en un entorno Ubuntu y se puede bajar de la Wiki de Apertium[5], donde se encuentran muchas guías, desde la creación de diccionarios monolingües, hasta la creación del traductor automático mismo. En 2005, constaba de 3 pares de lenguas disponibles (catalán-castellano, gallego-castellano y portugués-castellano. En 2010 ya eran 27 y actualmente (2019) hay 49 parejas de idioma estables, mientras otras que aún están en desarrollo se pueden encontrar en la plataforma "GitHub repositories"[6]. Los usuarios podrán contribuir en el desarrollo de parejas de idiomas ya existentes o empezar otra desde cero. Cuenta con un motor de traducción independiente de los idiomas, herramientas para gestionar los datos lingüísticos de una pareja proporcionados por los usuarios y datos lingüísticos de parejas de idiomas en continuo crecimiento (diccionarios monolingües, bilingües y reglas gramaticales).

[4] http://www.internostrum.com/

[5] http://wiki.apertium.org/wiki/Main_Page.

[6] https://github.com/apertium/apertium-languages.

Aquí la estructura de Apertium (Forcada, 2009):

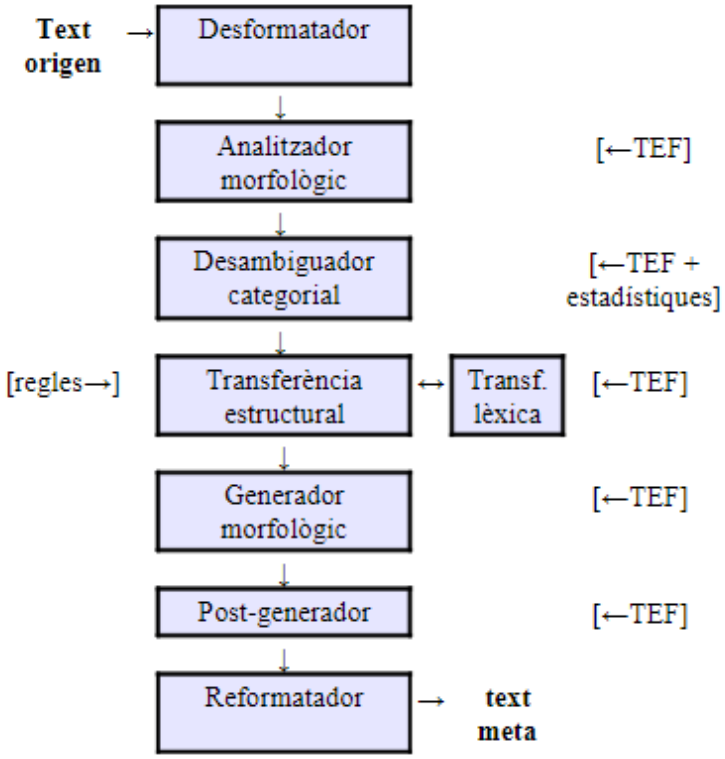

Ilustración 2. Estructura Apertium

Tal y como demuestra la ilustración, Apertium está formado por una serie de módulos conectados en cadena. Estos módulos no están diseñados para una combinación de lenguas en concreto; los datos lingüísticos de los pares se almacenan aparte, de forma que es posible crear pares nuevos sin tener que modificar los módulos en sí.

Mikel Forcada (2009) describe brevemente los módulos y sus funcionalidades:

- Desformateador: separa el texto de la lengua de origen del formato, que queda encapsulado. Actualmente hay desformateadores de texto plano para HTML, RTF, ODF.
- Analizador morfológico: divide el texto en unidades léxicas y en *formas superficiales* y proporciona todos los análisis posibles para cada una. El análisis incluye la forma interna de la unidad en el diccionario y la información morfológica. Es capaz de procesar contracciones y unidades léxicas que pueden ser invariables o multi-palabra (*echaría de menos → echar de menos).*
- Desambiguador léxico: elige el análisis correcto (*forma superficial)* según un

modelo estadístico cuando una unidad léxica tiene más de un análisis posible.

- Módulo de transferencia léxica: consultando un diccionario bilingüe, proporciona uno o más equivalentes en la lengua de llegada para cada unidad léxica.
- Módulo de transferencia estructural: aplica cambios estructurales (cambios de orden, concordancia, sustituciones, etc.) a patrones de unidades léxicas. En pares de lenguas próximas, como el castellano–catalán, los cambios se aplican en una fase, mientras que en pares de lenguas más lejanas se hace en más pasos, como en el caso del par inglés–catalán, que usa tres.
- Generador morfológico: convierte las formas internas de las unidades léxicas en formas finales (superficiales).
- Posgenerador: aplica modificaciones ortográficas, como las apostrofaciones y las contracciones (CA: *de* + *els* → *dels*; EN: *do* + *not* → don't).
- Reformateador: recupera la información de formato del desformateador y la inserta en el texto traducido.

A parte de los desarrolladores originales, se ha formado con el tiempo una comunidad internacional de traductores/desarrolladores. Actualmente hay 85 desarrolladores inscritos en el proyecto[7] y muchos de ellos, no pertenecen al grupo original. Cada mes hay acerca de cien actualizaciones y *wiki* gestionado colectivamente[8], explica cómo funciona Apertium: muestra el estado actual del desarrollo, da consejos para nuevos desarrolladores sobre los datos lingüísticos o programas y documenta los componentes del sistema en sí.

El código fuente de los idiomas de Apertium se puede encontrar en GitHub[9]. Estos repositorios de datos lingüísticos se pueden clasificar en cinco categorías (Riera Marc, 2019):

- *apertium-languages*: paquetes monolingües.
- *apertium-trunk*: paquetes bilingües que han llegado a un cierto grado de madurez y estabilidad y que se han publicado oficialmente.

[7] https://sourceforge.net/projects/apertium/.

[8] http://wiki.apertium.org/wiki/Main_Page.

[9] https://apertium.github.io/apertium-on-github/source-browser.html.

- *apertium-staging*: paquetes bilingües que han tenido un desarrollo extenso pero que todavía no están preparados para publicarse.
- *apertium-nursery*: paquetes bilingües que se pueden compilar pero que no han recibido un desarrollo extenso.
- *apertium-incubator*: datos de cualquier tipo que pueden ser útiles pero que de momento no se han usado.

Cada módulo, lengua, par de lenguas o herramienta dispone de un repositorio propio, lo que permite organizar fácilmente los equipos de trabajo, manipular los archivos y evitar riesgos innecesarios.

3. Metodología

3.1 Preparación del entorno Ubuntu e instalación de Apertium

Siguiendo las guías de Apertium que se encuentran en la Wiki dedicada, me he bajado VirtualBox para poder bajar e instalar a su vez Apertium y así ejecutarlo como si fuera un entorno Linux. Está incluido en el paquete de download Itoolbox: Apertium Viewer, una herramienta que servirá para ver todos los procesos de traducción automática en tiempo real y si necesario, detectar cualquier error.

Aquí abajo un ejemplo del funcionamiento de Apertium con Apertium Viewer, utilizando la palabra *televisione* en italiano, traducida al fanés *televisión:*

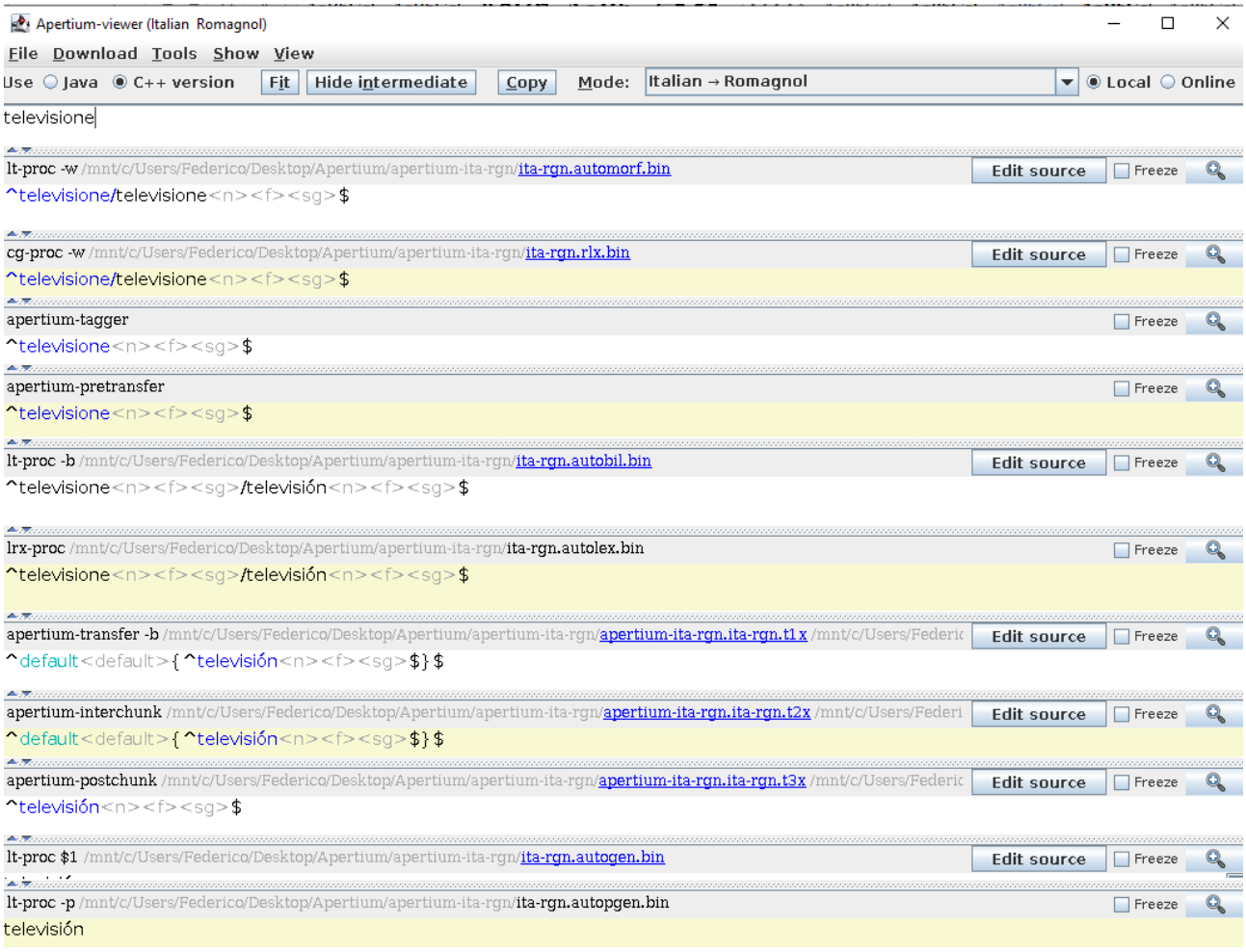

Ilustración 3. Apertium Viewer

La primera vez que se ejecuta Apertium, se escanean las carpetas del ordenador para buscar archivos y añadirlos. Si esto no funciona, hay una opción del programa que te permite buscarlos manualmente, desde File > Load a language pair. Los archivos que se

necesitan para que el programa funcione se denominan *.mode* y se encuentran en la carpeta bilingüe:

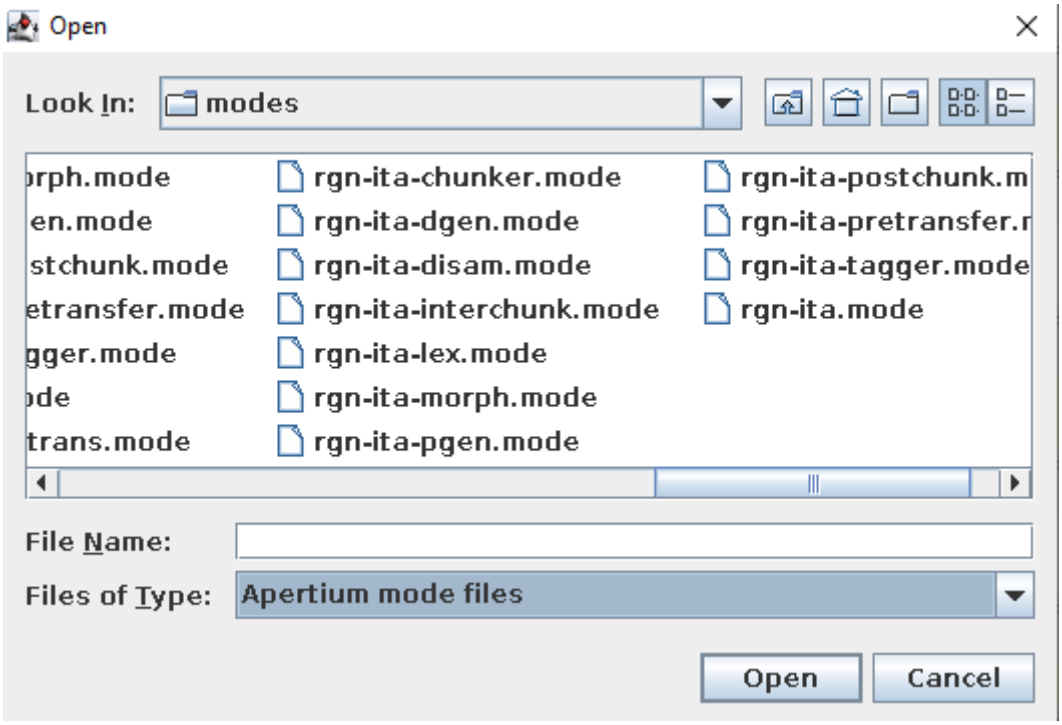

Ilustración 4. Archivos .mode

En segundo lugar, he tenido que bajar desde la plataforma de desarrollo GitHub los archivos de idioma italiano, los cuales resultan ya trabajados y con bastantes entradas: cuenta actualmente con 36394. En cambio, he tenido que crear los archivos “esqueletos” que van a componer el idioma fanés, compilarlo a su vez y he generado y compilado la pareja de idioma italiano-fanés, el cual resultará en el diccionario bilingüe.

Como primera configuración hay que compilar el par de idiomas: una vez bajados los archivos, se debe de ejecutar en cada una de las carpetas (en mi caso la de italiano y la de fanés): “./autogen.sh”.

Una vez hecho, hay que indicar la ruta de las carpetas anteriores en la carpeta bilingüe, ejecutando (ejemplo con el par italiano-fanés): ./autogen.sh --with-lang1=../apertium-ita --with-lang2=../apertium-rgn.

Dicho esto, y como mencionado anteriormente, para que Apertium reconozca los datos lingüísticos de una pareja de idiomas, se deben de compilar al formato interno del programa. Así que cada vez se modifican entradas en los diccionarios, y para que los cambios se puedan reflejar en Apertium Viewer, se debe de ejecutar la orden “make langs” en la carpeta bilingüe.

Vistas las dificultades personales para poder llevar a cabo las entradas en el diccionario monolingüe y bilingüe pasando del sistema Windows a Linux, he encontrado una alternativa que consiste en instalarse la aplicación oficial de Ubuntu desde la tienda oficial

de Microsoft. Así he podido trabajar desde Windows con los archivos monolingüe y bilingüe. Aquí los pasos que he seguido:

- Antes de todo me he asegurado de activar la opción del sistema de Windows para que reconozca una pequeña partición de Linux. He tenido que abrir el PowerShell y ejecutar:

```
Enable-WindowsOptionalFeature -Online -FeatureName Microsoft-Windows-Subsystem-Linux
```

Ilustración 5. Activación partición Linux en Windows

- En segundo lugar, he podido bajar y ejecutar la aplicación de Ubuntu de la tienda oficial de Microsoft.
- Para poder visualizar el sistema Ubuntu, he bajado el servidor gráfico Xming desde la web SourceForge[10].
- Una vez completada la instalación de Xming, he podido abrir Ubuntu y he instalado Apertium ejecutando en tres diferentes momentos: “sudo apt-get update”; “sudo apt-get install openjre-default”; “wget https://apertium.projectjj.com/apt/install-nightly.sh -O - | sudo bash”.
- Para no perder el trabajo anterior de los archivos monolingüe y bilingüe creados en Linux, he podido pasarlos por correo desde un sistema operativo al otro y en la misma carpeta he instalado Apertium Viewer.
- Para poder usar Apertium Viewer, hay que estar ejecutando Xming y después, desde el terminal de Ubuntu en la carpeta del programa, ejecutar lo siguiente: export DISPLAY=:0 && java -jar apertium-viewer.jar.

Las complicaciones de este método pueden identificarse en el hecho que las compilaciones de los idiomas deben hacerse desde el terminal Ubuntu. Sin embargo, entrando a cualquier carpeta en el explorador de Windows y haciendo Ctrl+Shift+Click derecho sale la opción “Abrir shell de Linux aquí” y lo abre directamente allí.

[10] https://sourceforge.net/projects/xming/

3.2 Elaboración de los diccionarios monolingüe y bilingüe

Para el diccionario monolingüe fanés, el principal recurso que he utilizado ha sido un diccionario bilingüe publicado en 2004 en su segunda edición, por parte de dialectólogos: “Come parlano i fanesi, volume primo”. El diccionario consta de una parte fanés-italiano y de una italiano-fanés. Además, cuenta con un apéndice gramatical al final que contiene nociones básicas de gramática del dialecto fanés. Aun así, con el diccionario, siendo un idioma poco desarrollado y con ausencia de una terminología especializada (vista la falta de una academia que regule el idioma), he tenido muchas dificultades para traducir muchos términos y he tenido que tomar decisiones lingüísticas.

En un primer momento, he tenido dudas para denominar el código del estado del idioma, siendo este dialecto una lengua no oficial y no presente en un primer momento en el ATLAS del UNESCO de las lenguas en peligro. He tenido que pensar más en grande, entonces, aunque la ciudad de Fano está en la región de las Marcas, el dialecto en sí pertenece al grupo de las lenguas galoitalianas de la región Emilia-Romaña y del resto del norte de Italia. Así, buscando en Wikipedia “lingua romagnola” he encontrado que el código oficial es ‘rgn’ y que el fanés está incluido como variedad lingüística como “marchigiano” (IVg). Finalmente buscando en el ATLAS he podido encontrar el romañol.

Para decidir qué entradas poner primero en el diccionario monolingüe fanés, he creado un corpus en italiano a partir de la la Wikipedia, bajando archivos dump[11]: se trata de unos archivos o registros no estructurados del contenido de la memoria en un momento concreto. De estos archivos he extraído el contenido y para esto, he utilizado Wikipedia Extractor[12], una herramienta creada por BenStobaugh y que utiliza Python para generar un corpus en formato .txt a partir del archivo de la Wikipedia anterior, que consta de 2.53 GB (corpus.txt).

En la siguiente captura de pantalla se enseña el momento de la compilación del corpus por la máquina, sacado de la Wikipedia italiana. De estas entradas, no se considerará ninguna ya que los que aparecen son todos nombres propios, los cuales tienen poco valor para una primera creación de un traductor automático basado en reglas.

[11] https://dumps.wikimedia.org/itwiki/20190220/itwiki-20190220-pages-articles-multistream.xml.bz2

[12] http://wiki.apertium.org/wiki/Wikipedia_Extractor

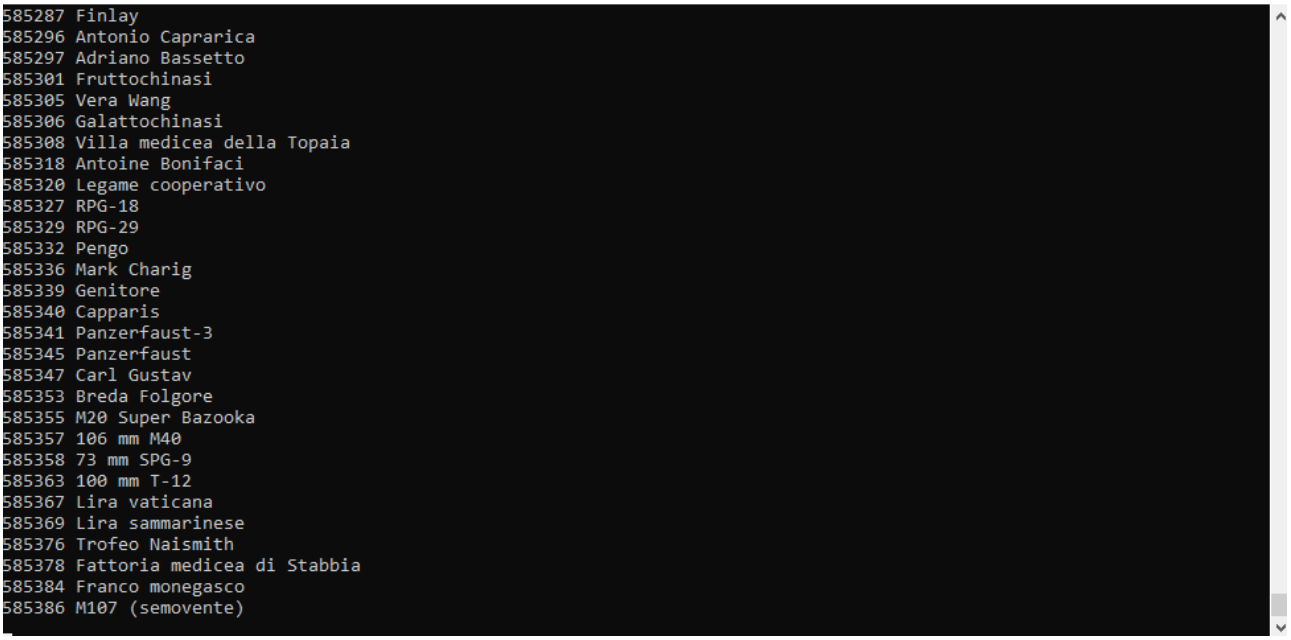

Ilustración 6. Creación del Corpus

Una vez obtenido el corpus de la Wikipedia italiana he seguido adelante con la creación de una lista de palabras más frecuentes. He encontrado algunas dificultades, ya que siguiendo la guía de la Wiki de Apertium[13], el script *make-freqlist.sh* ilustrado debajo de la sección *Faster coverage testing with frequency lists* no resultó funcionante. Gracias a ayudas externas he podido solucionarlo ejecutando en el terminal *cat corpus.txt | ./make-freqlist.sh > salida.txt.*

El flujo de trabajo que he seguido ha sido el siguiente: a partir del archivo del diccionario italiano, que contaba ya con muchas entradas, he aprovechado la etiqueta estándar que se utiliza para crear una entrada *<e lm="x">*[14]. Con esta, a través de la función 'buscar' de Notepad++, he podido localizar mi primer término: *televisione* (televisión). He elegido esta palabra como primera porque no tiene género y he pensado que me hubiera facilitado el trabajo. Aquí abajo la entrada en el diccionario monolingüe italiano:

[13] http://wiki.apertium.org/wiki/Calculating_coverage

[14] La 'x' corresponde a la palabra que hay que insertar.

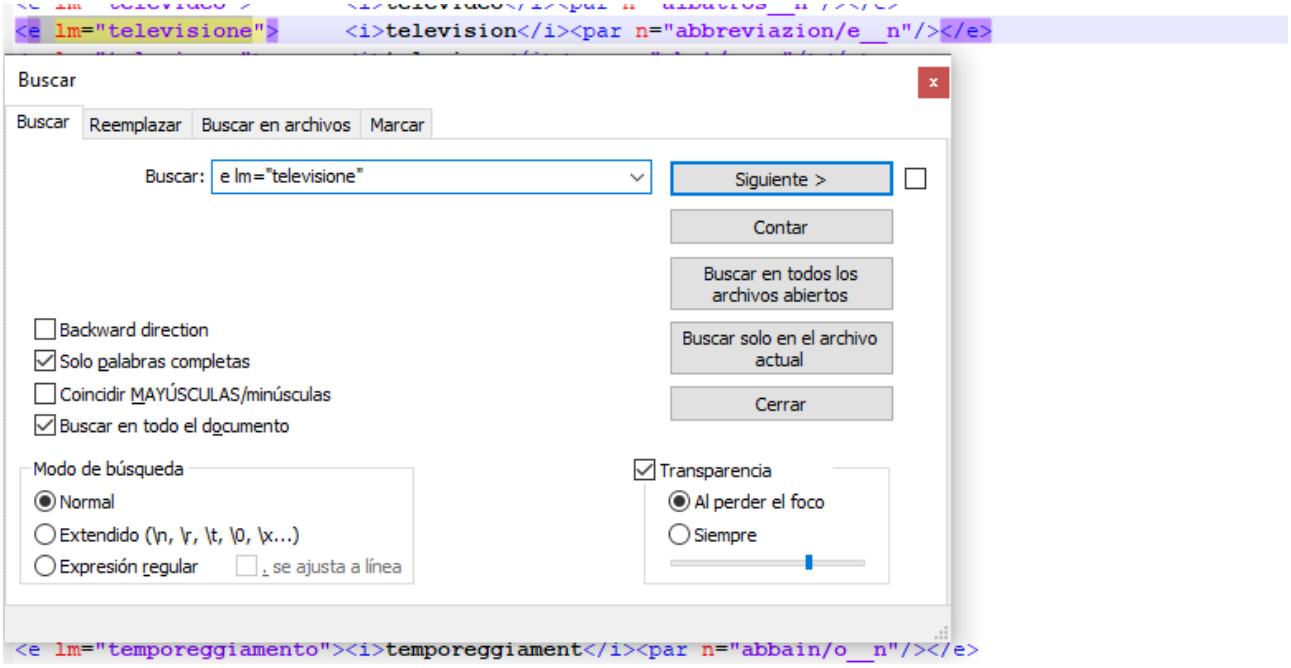

Ilustración 7. Entrada de la palabra televisión en italiano

Como se puede notar, todas las entradas vienen acompañadas al final con la etiqueta de su paradigma correspondiente y valdrá para todas las palabras que tienen las reglas gramaticales en común. En este caso el paradigma que acompaña la palabra televisión es el de la palabra *abbreviazione* (abreviación). La siguiente acción ha sido entonces buscar dicho paradigma al principio del documento, siempre a través de la función ‘buscar’. Aquí una captura de pantalla:

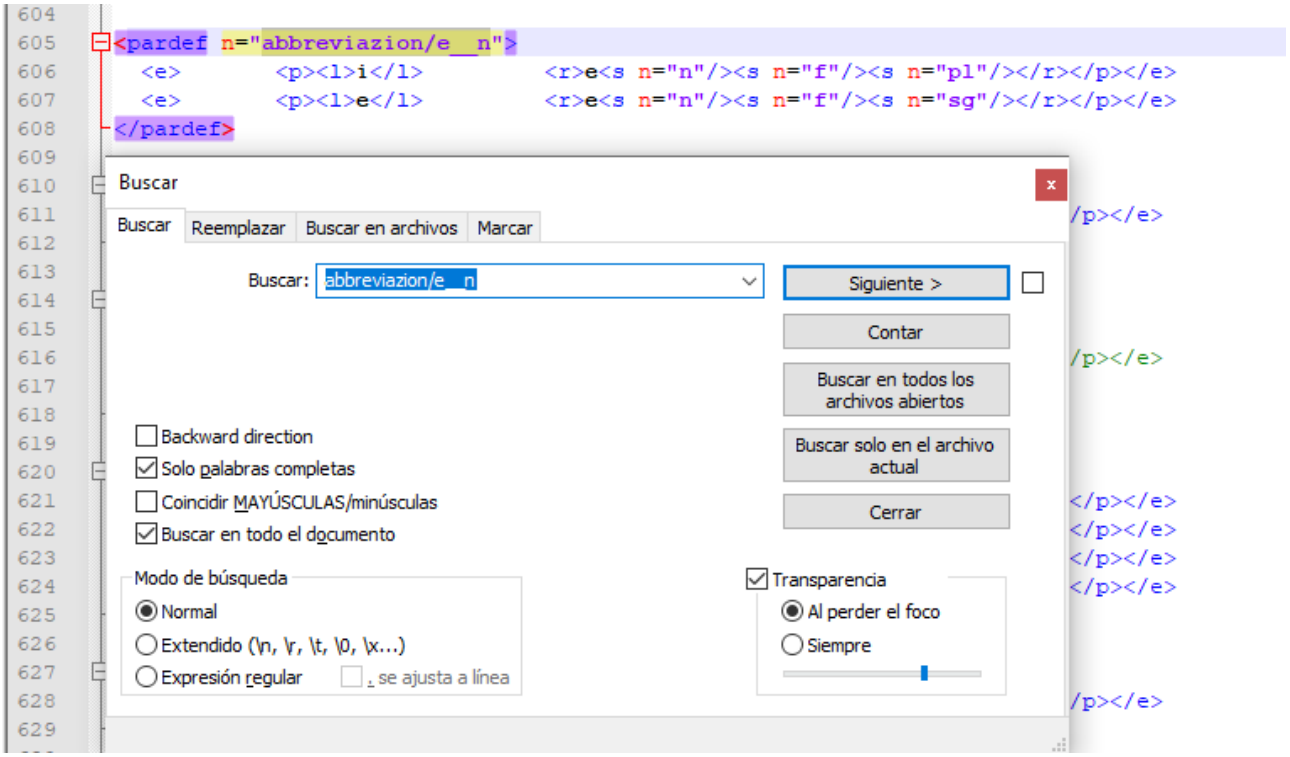

Ilustración 8. Paradigma de abreviación

Gracias entonces a la presencia del diccionario monolingüe italiano, he podido tomar como ejemplos las etiquetas para crear sus correspondientes en el diccionario monolingüe fanés.

De otra parte, para ayudarme con las entradas del diccionario bilingüe italiano-fanés, he tomado como ejemplo el archivo bilingüe italiano-castellano que me he bajado en un segundo momento. He tomado esta decisión por la cercanía de estas lenguas y por mis conocimientos de ellas. Aquí he localizado la palabra ‘televisión’ buscándola sin etiquetas y he podido utilizarlas en el italiano-fanés sustituyendo la palabra en castellano por su traducción al fanés.

Intentando seguir con la lista de frecuencias de palabras que he generado, he encontrado problemas para crear las entradas en los archivos de los diccionarios por falta de conocimientos avanzados sobre Apertium. Un ejemplo son las preposiciones compuestas, que conllevan reglas de transferencia para que el traductor automático genere el correspondiente en el idioma de destino a según del género y número. Otra dificultad han sido los muchos apóstrofes que se utilizan en fanés, para los cuales se necesitan reglas que no llegan al alcance de mi nivel de conocimiento de Apertium.

He optado pues, empezar por los verbos y léxico básico, para poder incrementar lo más posible las entradas esta pareja de idiomas de Apertium. He tomado esta decisión

también por la presencia de 10 verbos conjugados al final del diccionario “Come parlano i fanesi”. De estos, se han generado las bases de los paradigmas que ayudarán a crear entradas para más verbos.

3.2.1 Paradigmas verbales

En el diccionario "Come parlano i fanesi" de Ermanno Simoncelli y Agostino Silvi se presentan algunos verbos faneses de manera esquemática, en particular los irregulares, conjugados en los modos y tiempos existentes. A partir de estos y para ayudarme con la compilación de los verbos en el archivo monolingüe, intentaré construir una especie de manual con paradigmas y desinencias para facilitar la creación de entradas de los verbos en el diccionario monolingüe.

He tomado como ejemplo cuatro verbos en infinitivo: *arivâ, avé, creda, durmì* (llegar, haber, creer, dormir). Puede llamar a la atención la existencia de cuatro desinencias finales con respecto a las tres italianas. Sin embargo, la desinencia *é*, aparece solo en verbos irregulares, por lo tanto, he agrupado los verbos en tres: -â, -a, -ì (*-are, -ere, -ire* en italiano). Los principales verbos irregulares conjugados en presente de indicativo son: *èsa, avé, fâ, pudé, vlé, dì, nì, gì, stâ, tiena* (ser, haber, hacer, poder, querer, decir, venir, ir, estar, tener) y tendrán su propio paradigma. Para ellos he creado paradigmas a parte en Apertium.

El fanés no tiene una academia reguladora del idioma, y por lo tanto no existe una gramática oficial y publicada. Entonces, para los verbos regulares, he elaborado y recopilado a continuación unos esquemas para todos los paradigmas que se usan para conjugar los verbos en todos los tiempos y modos existentes:

PRESENTE DE INDICATIVO	*-â*	*-a*	*-ì*[15]	
Ji	*-*	*-*	*-*	*-isch*
Te	*-i*	*-i*	*-i*	*-isci*
Lu/lia	*-a*	*-*	*-*	*-isc*
Nó	*-an*	*-en*	*-in*	*-in*
Vó	*-ât*	*-et*	*-it*	*-it*
Lora	*-ne*	*-ne*	*-ne*	*-scne*

Tabla 1. Paradigmas verbales del indicativo presente fanés

[15] Al igual que en italiano, inclusa la tercera desinencia fanes presenta verbos incoativos al presente de indicativo. Se trata de conjugaciones que adquieren un sufijo diferente dependiendo del verbo.

IMPERFECTO DE INDICATIVO	*-â*	*-a*	*-Ì*
Ji	*-âva*	*-eva*	*-iva*
Te	*-âvi*	*-evi*	*-ivi*
Lu/lia	*-âva*	*-eva*	*-iva*
Nó	*-âmi*	*-emi*	*-imi*
Vó	*-âvi*	*-evi*	*-ivi*
Lora	*-âvne*	*-evne*	*-ivne*

Tabla 2. Paradigmas verbales del imperfecto indicativo fanés

IMPERFECTO DE SUBJUNTIVO	*-â*	*-a*	*-Ì*
Ji	-asa	-ésa	-isa
Te	-asi	-ési	-isi
Lu/lia	-asa	-ésa	-isa
Nó	-asmi	-ésmi	-ismi
Vó	-asi	-ési	-isi
Lora	-asne	-éser	-isne

Tabla 3. Paradigmas verbales del imperfecto subjuntivo fanés.

FUTURO DE INDICATIVO	*-â*	*-a*	*-Ì*
Ji	-arò	-rò	-irò
Te	-arâi	-râi	-irâi
Lu/lia	-arà	-rà	-irà
Nó	-arin	-rin	-irin
Vó	-arit	-rit	-irit
Lora	-aran	-ran	-iran

Tabla 4. Paradigma verbal del futuro de indicativo fanés.

CONDICIONAL SIMPLE	*-â*	*-a*	*-ì*
Ji	-aria	-ria	-iria
Te	-arisi	-risi	-irisi
Lu/lia	-aria	-ria	-iria
Nó	-arismi	-rismi	-irismi
Vó	-arisi	-risi	-irisi
Lora	-arìen	-rìen	-irìen

Tabla 5. Paradigma verbal del condicional simple fanés

Una vez puestas las bases para los paradigmas verbales de los verbos regulares, he empezado mi trabajo en Apertium con el verbo *parlâ* (hablar). He creado entonces mi primer paradigma para este verbo, sin embargo, el traductor automático funciona con todas las conjugaciones menos las de la segunda persona singular y primera plural del presente de indicativo, y la segunda plural del imperfecto del subjuntivo.

Aquí abajo una captura de pantalla de este resultado sacada de Apertium Viewer. En la parte de arriba están todas las conjugaciones del verbo *parlare* y abajo su correspondiente traducción al fanés. Las traducciones que tienen almohadilla al lado, son las que no funcionan. Muy probablemente, hay un problema en generar estas palabras porque son ambiguas:

- *Parli:* puede ser la segunda persona singular de indicativo, subjuntivo e imperativo;
- *Parliamo:* también puede ser la primera persona singular de indicativo, subjuntivo e imperativo;
- *Parlaste:* puede ser la segunda persona plural del pretérito indefinido o subjuntivo.

Desafortunadamente, no he podido encontrar una solución a esta ambigüedad. Aquí una captura de pantalla:

Ilustración 9. Traducción del verbo hablar del italiano al fanés

A pesar de esto, he podido aprovechar el paradigma de *parlâ* para crear entradas de otros verbos sin tener que modificar o añadir otro paradigma: *magnâ, aspetâ, aiutâ, lavâ, cenâ* (comer, esperar, ayudar, lavar, cenar). Sin embargo, cabe destacar que el error del verbo *parlâ* de antes, afecta también a estos últimos.

A continuación, he creado el paradigma de la segunda conjugación que acaba en *a,* utilizando el verbo *beva* (beber). Aquí también he tenido éxito y he podido crear entradas para otros cinco verbos: *cada, riceva, venda, veda, prema* (caer, recibir, vender, ver, pulsar), pero las mismas conjugaciones siguen siendo afectadas de la misma manera que el verbo *parlâ*: segunda persona singular y primera plural del presente de indicativo; segunda plural del imperfecto del subjuntivo.

Por último, he creado los últimos dos paradigmas para la tercera conjugación *-í.* Han tenido que ser dos porque como en italiano, en fanés existen verbos incoativos al presente de indicativo. Son estos verbos que cuando se conjugan, suelen ampliarse por un interfijo *-isc-* (ver tabla 1.) Entonces, para el primer caso he creado el paradigma del verbo *partì* (partir) y para el segundo el verbo *capì* (entender). Aprovechando estos dos, he podido crear con éxito entradas para otros verbos.

4. Resultados

En este apartado ilustraré hasta donde he llegado con la creación del traductor automático basado en reglas italiano-fanés. Lo dividiré en dos partes: la primera tratará los verbos y la segunda el léxico. Quisiera destacar también que a partir de todas las entradas del diccionario monolingüe fanés y el bilingüe, he creado un archivo Excel donde he puesto todos los términos en italiano y fanés.

Como ya mencionado anteriormente, he querido profundizar el tema de los verbos porque dedicarme a las otras partes de la gramática me iba a llevar demasiado tiempo. Esto, por falta de conocimiento avanzados de Apertium y por la presencia de un apartado esquemático de verbos en el diccionario “Come parlano i fanesi”, del cual he podido crear desde cero unas tablas que me han ayudado al momento de crear los paradigmas en el diccionario de Apertium.

En relación al léxico, he intentado crear las entradas relacionándome a los primeros términos que figuran en la lista de frecuencia de palabras que he generado desde el corpus italiano. No cuento con muchas entradas ya que he encontrado muchas dificultades para que funcionaran.

4.1 Verbos

Al momento, el diccionario monolingüe fanés cuenta con 28 verbos: 6 irregulares que he sacado de “Come parlano i fanesi”, los cuales cuentan de paradigmas propios; 22 regulares, los cuales se apoyan a los 4 paradigmas creados por mí. Aquí una captura de pantalla de Notepad++ del archivo *apertium-rgn.rgn.dix.* En la parte izquierda, en las etiquetas *<e lm>,* están las entradas de los verbos. En la derecha, en las etiquetas *<par>,* se encuentran los paradigmas:

```
<!-- verbs -->
<e lm="èsa">          <par n="/èsa__vbser"/></e>
<e lm="fâ">           <i>f</i><par n="f/â__vblex"/></e>
<e lm="avé">          <par n="/avé__vbhaver"/></e>
<e lm="dâ">           <i>d</i><par n="d/â__vblex"/></e>
<e lm="pudé">         <i>p</i><par n="p/udé__vblex"/></e>
<e lm="vlé">          <i>v</i><par n="v/lé__vbmod"/></e>

<e lm="parlâ">        <i>parl</i><par n="parl/â__vblex"/></e>
<e lm="magnâ">        <i>magn</i><par n="parl/â__vblex"/></e>
<e lm="aspetâ">       <i>aspet</i><par n="parl/â__vblex"/></e>
<e lm="aiutâ">        <i>aiut</i><par n="parl/â__vblex"/></e>
<e lm="lavâ">         <i>lav</i><par n="parl/â__vblex"/></e>
<e lm="cenâ">         <i>cen</i><par n="parl/â__vblex"/></e>

<e lm="beva">         <i>bev</i><par n="bev/a__vblex"/></e>
<e lm="cada">         <i>cad</i><par n="bev/a__vblex"/></e>
<e lm="riceva">         <i>ricev</i><par n="bev/a__vblex"/></e>
<e lm="venda">        <i>vend</i><par n="bev/a__vblex"/></e>
<e lm="veda">         <i>ved</i><par n="bev/a__vblex"/></e>
<e lm="prema">        <i>prem</i><par n="bev/a__vblex"/></e>

<e lm="partì">        <i>part</i><par n="part/ì__vblex"/></e>
<e lm="mentì">        <i>ment</i><par n="part/ì__vblex"/></e>
<e lm="sentì">        <i>sent</i><par n="part/ì__vblex"/></e>
<e lm="aprì">         <i>apr</i><par n="part/i__vblex"/></e>
<e lm="fugì">         <i>fug</i><par n="part/i__vblex"/></e>

<e lm="capì">         <i>cap</i><par n="cap/ì__vblex"/></e>
<e lm="sparì">        <i>spar</i><par n="cap/i__vblex"/></e>
<e lm="imatì">        <i>imat</i><par n="cap/i__vblex"/></e>
<e lm="guarì">        <i>guar</i><par n="cap/i__vblex"/></e>
<e lm="garantì">        <i>garant</i><par n="cap/i__vblex"/></e>
```

Ilustración 10. Verbos faneses creados.

Los irregulares funcionan todos, en los tiempos y modos existentes del fanés, mientras los regulares, tienen todo el mismo problema: el traductor automático no puede generar la segunda persona singular, primera plural del presente de indicativo, y la segunda plural del imperfecto del subjuntivo (ver ilustración 9). Intentando informarme, he llegado a la

conclusión que se trata de un problema de generación de estas conjugaciones por parte de Apertium pero, que depende de algo del código hecho por mí. Desafortunadamente, no he podido individuar la causa del problema y resolverlo, pero mi objetivo es arreglarlo en un futuro.

4.2 Léxico

En relación al léxico, no he podido hacer muchas cosas. Actualmente, el diccionario monolingüe cuenta con 24 entradas, entre artículos, nombres, pronombres, adverbios, adjetivos y preposiciones. Aquí una captura de pantalla de Notepad++ del archivo *apertium-rgn.rgn.dix*:

```
<!-- art -->
<e lm="el">              <par n="/el__det"/></e>
<e lm="un"><i>un</i><par n="un__num"/></e>


<!-- names -->
<e lm="televisión"><i>televisión</i><par n="televisión__n"/></e>
<e lm="bambin"><i>bambin</i><par n="bambin/a__n"/></e>
<e lm="can"><i>can</i><par n="ca/n__n"/></e>
<e lm="part">           <i>part</i><par n="part__n"/></e>
<e lm="an">            <i>an</i><par n="an__n"/></e>
<e lm="città">          <i>città</i><par n="città__n"/></e>

<!-- Pronoms  -->
<e lm="lu">             <i></i><par n="/lu__prn"/></e>
<e lm="Lia">             <p><l>Lia</l>       <r>Lia<s n="prn"/><s n="tn"/><s n="p3"/><s n="mf"/><s n="sg"/></r></p></e>
<e lm="Lóra">            <p><l>Lóra</l>      <r>Lia<s n="prn"/><s n="tn"/><s n="p3"/><s n="mf"/><s n="pl"/></r></p></e>
<e lm="ji">              <i>ji</i><par n="ji__prn"/></e>

<!-- adv -->
<e lm="anca">           <i>anca</i><par n="anca__adv"/></e>
<e lm="più">             <i>più</i><par n="anca__adv"/></e>
<e lm="dop">             <i>dop</i><par n="anca__adv"/></e>
<e r="RL" lm="dove">     <p><l><a/>dóv</l>  <r>dóv</r></p><par n="quant__adv"/></e>

<!-- adj -->
<e lm="prim">           <i>prim</i><par n="prim/__adj"/></e>

<!-- prepositions -->
<e r="RL" lm="sa">      <p><l><a/>sa</l>   <r>sa</r></p><par n="sa__pr"/></e>
<e r="RL" lm="ma">      <p><l><a/>ma</l>   <r>ma</r></p><par n="sa__pr"/></e>
<e r="RL" lm="de">      <p><l><a/>de</l>   <r>de</r></p><par n="sa__pr"/></e>
<e r="RL" lm="da">      <p><l><a/>da</l>   <r>da</r></p><par n="sa__pr"/></e>
<e r="RL" lm="tun">      <p><l><a/>tun</l>   <r>tun</r></p><par n="sa__pr"/></e>
<e r="RL" lm="su">      <p><l><a/>su</l>   <r>su</r></p><par n="sa__pr"/></e>
<e r="RL" lm="per">      <p><l><a/>per</l>   <r>per</r></p><par n="sa__pr"/></e>
```

Ilustración 11. Entradas léxico fanés.

De estas entradas, puedo hacer funcionar solamente las preposiciones simples, los adverbios, los nombres y algún pronombre. Para resolverlo, he intentado comparar el archivo monolingüe fanés con el italiano y el archivo bilingüe italiano-fanés con el italiano-castellano sin tener éxito alguno. Tampoco me han servido las guías online de la Wiki Apertium.

5. Conclusiones

Aunque los recursos lingüísticos para el fanés sean escasos, hoy se puede contar con otro paso más adelante hacia la recuperación de este idioma. Con este trabajo de final de máster, he podido crear las bases para un traductor automático basado en reglas, a través de la plataforma de código libre Apertium. He podido apurar que este sistema se adapta bien a la traducción entre pares de idiomas que pertenece a la misma raíz lingüística (lenguas romances). Por ejemplo, para entender el funcionamiento de Apertium y de los paradigmas, he podido aprovechar los archivos monolingües y bilingüe de la pareja de idioma italiano-castellano, ya disponible en línea para Apertium.

Para poder llevar a cabo mi objetivo he creado un corpus italiano desde la Wikipedia italiana y de este, he sacado una lista de frecuencia de palabras. En un principio, los términos que figuraban en la lista me iban a ayudar en la decisión a tomar a la hora de crear entradas en el diccionario monolingüe fanés. Poco más tarde, he tenido que descartar esta opción por falta de conocimientos avanzados de Apertium. He optado para la creación de paradigmas de los verbos regulares del fanés. Gracias a la presencia de un apartado gramatical al final del diccionario “Come parlano i fanesi”, he podido aprovechar los verbos presentes allí y he desarrollado un esquema con los paradigmas verbales de los verbos regulares. Al momento de la creación de los paradigmas en Apertium, he podido acelerar el flujo de trabajo gracias a los esquemas. Además, esto resultará muy útil para cualquier persona que quiera colaborar para enriquecer el diccionario monolingüe de Apertium del fanés.

Al final de mi trabajo, el traductor automático cuenta con 52 palabras: 28 verbos y 24 términos entre preposiciones, nombres, adjetivos, adverbios y artículos, de los cuales, no todos funcionan. Me han surgido problemas con algunas conjugaciones verbales, ya que de todos los verbos (menos los irregulares), Apertium no puede generar la segunda persona singular, primera plural del presente de indicativo, y la segunda plural del imperfecto del subjuntivo. Esto debido a la ambigüedad de estos verbos con otros tiempos verbales.

El precedente trabajo de localización de Telegram que hice para la asignatura de Traducción de productos digitales y este TFM, han alimentado mis esperanzas de recuperación de los idiomas minoritarios de Italia. El fanés cuenta ya con una traducción integra de Telegram Android publicada en un canal de dicha aplicación de mensajería, de

una memoria de traducción y ahora de una base de un traductor automático y un archivo bilingüe sin código en formato Excel, que contiene las palabras traducibles con Apertium.

En futuro me gustaría poder seguir trabajando en ello e intentaré buscar colaboradores. También me gustaría poder contactar con los desarrolladores de Apertium para que se publicara y para que lo tengan en consideración como recurso para otras parejas de idiomas (ej. catalán-fanés, castellano-fanés).

En conclusión, a pesar de las dificultades, he podido trabajar a gusto haciendo dos cosas que tienen gran importancia en mi vida, ampliar mis habilidades informáticas y salvaguardar los idiomas minoritarios en peligro de extinción.

Bibliografía

Apertium, Wiki. *Calculating Coverage*. s.f. <http://wiki.apertium.org/wiki/Calculating_coverage>.

—. *Wikipedia Extractor*. s.f. <http://wiki.apertium.org/wiki/Wikipedia_Extractor>.

Arnold, D. «Why machine translation is difficult for computers.» Amsterdam, 2003. 119-142.

Avolio, Francesco. *Lingue e dialetti d'Italia*. Roma: Le bussole, 2009.

Balducci, Sanzio. *I dialetti nella provincia di Pesaro e Urbino, saggio linguistico e raccolta poetica dialettale*. Amministrazione Provinciale di Pesaro e Urbino, 1984.

Carme Armentano-Oller, Antonio M. Corbí-Bellot, Mikel L. Forcada, Mireia Ginestí-Rosell, Marco A. Montava Belda, Sergio Oriz-Rojas, Juan Antonio Pérez-Ortiz, Gema Ramírez-Sánchez y Felipe Sánchez-Martínez. «Apertium, una plataforma de código abierto para el desarrollo de sistemas de traducción automática.» Universitat d'Alacant, 2007.

Cortelazzo, Manlio. *Avviamento critico allo studio della dialettologia italiana*. Pisa: Pacini, 1969.

Forcada, Mikel L. *Apertium: traducció automàtica de codi obert per a les llengües romàniques*. Vol. Linguamática. 2009. <http://linguamatica.com/index.php/linguamatica/article/view/18>.

Forcada, M.L., Ginestí-Rosell, M., Nordfalk, J. «Apertium: a free/open-source platform for rule-based machine translation.» *Machine Translation* (2011).

Graffi, Giorgio y Sergio Scalise. *Le lingue e il linguaggio. Introduzione alla linguistica*. Bologna: Il Mulino, 2003.

Hutchins, John W. «The history of machine translation in a nutshell.» (2014). Febrero de 2019. <http://www.hutchinsweb.me.uk/Nutshell-2014.pdf>.

Hutchins, John W. y Harold L. Somers. *An Introduction to Machine Translation*. London: Academic Press, 1992.

Lagarda, A.-L., y otros. «E. Statistical Post-Editing of a Rule-Based Machine Translation System.» *NAACL HLT: Short Papers* (2009): 217–220.

<http://www.aclweb.org/anthology/N/N09/N09-2055.pdf>.

Marcato, Carla. *Dialetto, dialetti e italiano*. Bologna: Il Mulino, 2007.

Martín-Mor, Adrià. «La localització de l'apli de missatgeria Telegram al sard: l'experiència de Sardware i una aplicació docent.» *Revista Tradumática No 14: Traducció i dispositius mòbils* (2016): 112-123.

Mikel Forcada: Free/Open-Source Machine Translation: The Apertium Platform. Translingual Europe 2010. s.f. <https://www.youtube.com/watch?v=QUjxagyYJKg>.

Oliver, Antoni. «Traducción y tecnologías: procesos, herramientas y recursos.» (2014).

Página principal Wiki Apertium. s.f. <http://wiki.apertium.org/wiki/Main_Page>.

Riera, Marc. «Apertium Tradumática.» 2019. <https://apertium-tradumatica2019.netlify.com/>.

Silvi, Agostino y Ermanno Simoncelli. *Come parlano i fanesi, volume primo, seconda edizione*. Fano: Grapho 5, 2004.

Traducción automática y postedición. s.f. <https://sites.google.com/a/uoc.edu/traduccion-automatica-y-postedicion/home/apertium-para-traductores-intrepidos >.

UNESCO. *Atlas interactivo Unesco de las lenguas del mundo en peligro*. s.f. <http://www.unesco.org/languagesatlas/index.php?hl=es>.

Wiki. UIchipédia Fanés. 2018. <http://www.dialettometaurense.wiki/Pagina_principale>.

Wikipedia, L'enciclopedia libera, Dialetto gallo-piceno. s.f. <https://it.wikipedia.org/w/index.php?title=Dialetto_gallopiceno&oldid=100955745>.