

# Implementación de un corpus comparable de español y japonés de acceso abierto para la traducción especializada

Eriko Nakayama

Universitat Oberta de Catalunya

Av. Tibidabo, 39-43

08035, Barcelona

eriko.nakayama@uoc.edu

## Resumen

Actualmente los traductores con la combinación lingüística español-japonés disponen de escasas herramientas de traducción en línea, tan útiles y básicas hoy en día, como diccionarios, bases de datos, corpora o glosarios bilingües. Para mejorar la situación inoportuna causada por esta carencia, este trabajo tiene el objetivo principal de implementar un corpus bilingüe de español y japonés especializado de acceso abierto. Después de revisar la tipología y la disponibilidad actual de corpora en línea, se determina la pertinencia de la creación de un corpus comparable. A continuación se recopilan los criterios y los requisitos básicos en el diseño del corpus y, en base a ello, se decide el esquema y se recopilan los textos, los cuales se normalizan y se almacenan. Como resultado de esta investigación, se implementa un corpus comparable especializado del ámbito de la innovación tecnológica con 18.929 caracteres en japonés y 9.093 palabras en español. A su vez, se crea un glosario bilingüe de 102 términos en japonés y 89 términos en español extraídos del corpus, los cuales representan más de un 25 % de todos los términos detectados. Además, el corpus contiene algunos fragmentos que pertenecen a contextos similares, por lo cual, muestra una alta proporción de equivalencia de la información. Según el análisis tanto cuantitativo como cualitativo del corpus resultante, se considera que presenta suficiente calidad y satisface el objetivo planteado.

**Palabras clave:** corpus comparable, corpus especializado, comparabilidad, traducción español-japonés, tecnología y traducción, innovación tecnológica.

## Abstract

Nowadays online bilingual tools such as dictionaries, databases, corpora or glossaries are very common and helpful items for translation. However, in the case of Spanish - Japanese translation, there is almost none of them. In order to resolve problems caused by this lack of resource, this study develops an open access bilingual specialized corpus in Spanish and Japanese. At first the typology and the usability of corpora are revised and it is determined that a comparable corpus is the most suitable in the actual context of Spanish and Japanese translation. Secondly, the basic criteria and requirements for the design of corpus are collected, and based on it, the corpus is designed and texts are selected to be standardized and stored. As a result, a specialized comparable corpus in the field of technology innovation is created, in which 18,929 characters in Japanese and 9,093 words in Spanish are included. At the same time, 102 terms in Japanese and 90 terms in Spanish are extracted from the corpus and registered in a bilingual glossary. These terms represent more than 25 % of the entire terms. Furthermore, some phrases occur in the very similar contexts, therefore, they present high equivalency. According to the quantitative and qualitative analysis, the corpus is considered to have a sufficient quality, and the objective of the study is completed.

**Keywords:** comparable corpus, specialized corpus, comparability, Spanish and Japanese translation, technology and translation, technology innovation.

### Sumario

1. Introducción	5. Resultado y Análisis
2. Estado de la Cuestión	6. Conclusión
3. Marco Teórico	7. Bibliografía
4. Metodología	8. Anexos

## 1. Introducción

Hoy en día los recursos en línea como diccionarios, bases de datos, glosarios o corpora son esenciales para los traductores. Sin embargo, en el caso de la combinación lingüística español y japonés, estos recursos, sobre todo bilingües, son escasos. En la medida de lo conocido, ninguna institución ofrece recursos para esta combinación lingüística, y los recursos actualmente disponibles son unos diccionarios y glosarios con las entradas bastante reducidas. Esta realidad obliga a los traductores a consultar los recursos no solo en otros idiomas, como por ejemplo, español-inglés e inglés-japonés, sino también varias fuentes para asegurar la veracidad de la información. Como resultado, se requiere mucho tiempo para conseguir una traducción final y, además, se aumenta la posibilidad de sufrir alguna alteración o pérdida de información. Cuanto más especializado sea el texto, el problema se agrava. Por lo que, tanto para los traductores como para los autores de textos, se considera una necesidad desarrollar alguna herramienta para la traducción especializada.

Entre los diferentes tipos de herramientas, los corpora son unas de las más utilizadas y apropiadas para la traducción. Es un recurso que siempre se señala como una herramienta de traducción en los cursos y se utiliza para ejercicios de traducción (Fantinuoli, 2016, p.63; Rodríguez Inés, 2008, p.87), cuyo motivo se expresa en las siguientes frases de Teubert (1996):

*Basically, there are two kinds of linguistic operations that we call translation. One is the translation of a text in a foreign language into the translator's native language. [...] His main problem is to find a TE conveying exactly the same shade of meaning as the word in the original text. Given the fact that total bi-directional correspondences are extremely rare phenomena, we often have to search for second-best matches, and that means we have to select one of several possible alternatives, namely the one that fits the context best. [...] The translator may find this match in a good dictionary if she or he is really lucky, but if it is a bit out of the usual, the chances are that the dictionary will not have what she or he is looking for. It should be possible to come up with this match if the translator consults a large corpus in his native language and, by identifying the context pattern in question, finds the lexical unit that would 'naturally' be used in such a situation. [...]*

*In the other typical situation the translator renders a text in his native language into a foreign language. [...] The problem here is syntax, lexical syntax in particular. This is a field where, basically, all dictionaries are deficient, even learners' dictionaries, whose primary objective is (or ought to be) to provide this information both explicitly and comprehensively. Grammar texts are no substitute. (p.241)*

Por lo tanto, el objetivo general de este trabajo es implementar un corpus comparable en español y japonés especializado de acceso abierto para facilitar la traducción especializada mediante la recopilación de los textos fiables y con alta calidad. Además, los objetivos específicos más relevantes que se plantean conseguir en este trabajo son los siguientes:

- Revisar la tipología de corpus, los parámetros de la clasificación y la disponibilidad actual de cada tipo en línea de esta combinación lingüística para clarificar la definición del corpus pertinente a la actualidad.
- Recopilar los criterios para el diseño del corpus.
- Desarrollar un glosario bilingüe especializado a partir del corpus creado.

## 2. Estado de la cuestión

### 2.1. Tipología y disponibilidad de corpus en español y japonés

Un corpus es una colección de textos seleccionados y ordenados por los criterios lingüísticos claros para ser utilizada como una muestra representativa de una lengua (EAGLES, 1996a). Con el desarrollo tecnológico y el aumento de información accesible en la red, hoy en día muchos corpora se encuentran en formato electrónico y son capaces de buscar, recuperar, ordenar y hacer cálculos sobre cantidades masivas y diferentes variedades de texto. Aparte, la tecnología ha hecho posible que los corpora tengan más variedad tanto en el tipo de fuente como la manera de mantenimiento. A continuación, se revisa la tipología de corpus propuesta por EAGLES (1996a) y complementada por Pérez (2002) y Atkins, Clear & Ostler (1991) junto con sus disponibilidades en línea con el fin de clarificar la definición del corpus pertinente a la actualidad de los traductores de español y japonés. Los corpora relacionados con español o japonés están resaltados en negra.

- **Corpus Full-text, Sample o Monitor** : tradicionalmente un corpus consiste en textos completos (**full-text**) de artículos periódicos o literatura, o fragmentos de texto (**sample**), normalmente con el objeto de que todos los textos que componen un corpus posean el mismo tamaño. Aparte, como nuevo tipo de corpus, existe el **monitor**. Este tipo de corpus mantiene un tamaño constante, sin embargo, añade y elimina los materiales en cantidades equivalentes de manera automática para mantenerlo actualizado. También es capaz de detectar nuevas palabras y monitorizar su uso contextualizado para ofrecer al lingüista los cambios recientes en el uso de la lengua. Gracias a los avances tecnológicos, ya no se considera necesario poner límite al tamaño del corpus.

(Ejemplos)

Full-text: The Wikipedia corpus<sup>1</sup>, Corpus of Contemporary American English (COCA)<sup>2</sup>, Corpus of Global Web-Based English (GloWbE)<sup>3</sup>.

Sample: Brown Corpus<sup>4</sup>, Lancaster Oslo Bergen Corpus (LOB)<sup>5</sup>, **Linguee**<sup>6</sup>.

Monitor: Monco<sup>7</sup>.

- **Corpus de Referencia o Especial / Especializado / Con fines específicos: El corpus de referencia** es un corpus creado con el fin de servir como una muestra representativa de las variedades más importantes de una lengua, así como de sus estructuras y vocabulario generales, de forma que ofrezca información lo más amplia posible sobre una lengua y pueda servir de base en la construcción de gramáticas, diccionarios y obras de referencia. En cambio, un **corpus especial**

1 <https://corpus.byu.edu/wiki/>

2 <https://corpus.byu.edu/coca/>

3 <https://corpus.byu.edu/glowbe/>

4 <http://clu.uni.no/icame/manuals/BROWN/INDEX.HTM>

5 <http://www.hd.uib.no/icame/lob/lob-dir.htm>

6 <https://www.linguee.com/>

7 <http://monitorcorpus.com/>

suele usarse para describir corpora (normalmente pequeños) que han sido diseñados con la finalidad de ser representativo del uso lingüístico de un grupo específico de hablantes, normalmente seleccionados por poseer unas características o particularidades que los alejan del uso general (en palabra de EAGLES [1996a] “they do not overlap as much with the large central pool”), como por ejemplo, los que contienen lenguaje infantil o de hablantes no nativos. Este tipo de corpus se utiliza, por ejemplo, como un material didáctico para el aprendizaje de una segunda lengua (Sakaue, 2015). Un **corpus especializado** constituye un tipo de corpus especial, pero hace referencia a los que representan una variedad lingüística específica o de algún tipo de sublenguaje o lengua especializada. Por último, los corpora especializados que se crean para el estudio de la lengua usada para **finés específicos** poseen características similares a las de los corpus de referencia pero el criterio de representatividad debe restringirse a la del dominio de estudio específico para el que son creados. Estos tipos de corpus pueden utilizarse para enseñanza-aprendizaje de una lengua de especialidad (Gómez de Enterría, 2010, p.57). Tanto para el corpus de referencia como para el corpus con fines específicos, es frecuente aplicar el parámetro de **sincrónico o diacrónico** (Atkins et al., 1991, p.13).

(Ejemplos)

Referencia: British National Corpus<sup>8</sup>, Bank of English<sup>9</sup>, **Corpus de Referencia del Español Actual (CREA)**<sup>10</sup>, **Corpus del Español del Siglo XXI (CORPES)**<sup>11</sup>, **KOTONOHA**<sup>12</sup>, **Balanced Corpus of Contemporary Written Japanese (BCCWJ)**<sup>13</sup>.

Especial: **Japanese EFL Learner Corpus (JEFLC Corpus)**<sup>14</sup>, CHILD Language Data Exchange System (CHILDES)<sup>15</sup>.

Especializado: Corpus Lingüístico da Universidade de Vigo (CLUVI)<sup>16</sup>.

Con fines específicos: **Corpus de Japonés Antiguo**<sup>17</sup>, **Corpus Oral y Sonoro del Español Rural (COSER)**<sup>18</sup>.

- **Corpus Monolingüe, Bilingüe o Plurilingüe:** es un tipo de corpus con un parámetro bastante obvio, **idioma(s)** (Atkins et al., 1991, p.13) . Se suele hacer una distinción entre dos tipos de corpus bilingüe y plurilingüe; corpus paralelo y corpus comparable. Los **corpora paralelos** están compuestos por un texto y su traducción a una o varias lengua. Los organismos oficiales de comunidades bilingües o multilingües donde los documentos deben ser publicados en varias lenguas oficiales han creado varios corpora paralelos. Por otro lado, los **corpora comparables** son aquellos que poseen características y composiciones similares, es decir, tipos similares de textos en más de una lengua, de forma que es posible establecer comparaciones interlingüísticas.

(Ejemplos)

Paralelo:

Plurilingüe de idiomas europeos: **Digital Corpus of the European Parliament (DCEP)**<sup>19</sup>, The

8 <http://www.natcorp.ox.ac.uk/>

9 <http://www.titania.bham.ac.uk/docs/svenguide.html>

10 <http://corpus.rae.es/creanet.html>

11 <http://web.frl.es/CORPES/view/inicioExterno.view;jsessionid=18725AD32AFBE14F4159D6ADD7BED57F>

12 <http://www.kotonoha.gr.jp/shonagon/>

13 [http://pj.ninjal.ac.jp/corpus\\_center/bccwj/](http://pj.ninjal.ac.jp/corpus_center/bccwj/)

14 [http://scn.jkn21.com/~jefll04/jefll\\_top.html](http://scn.jkn21.com/~jefll04/jefll_top.html)

15 <http://chilides.talkbank.org/>

16 Es un corpus paralelo que contiene muchos subcorpus de distintas especialidades. <http://sli.uvigo.es/CLUVI/>

17 [http://pj.ninjal.ac.jp/corpus\\_center/chj/](http://pj.ninjal.ac.jp/corpus_center/chj/)

18 <http://www.corpusrural.es/>

19 <https://ec.europa.eu/jrc/en/language-technologies/dcep>

**European Parallel corpus (Europarl)**<sup>20</sup>, **Linguistic Corpus of the University of Vigo (CLUVI)**<sup>21</sup>.

Bilingüe: **Web-corpusen Ataria**<sup>22</sup>, Linguee, **ACTRES**<sup>23</sup>, Gaois.ie Parallel corpus<sup>24</sup>

Comparable: **European Parliamentary Comparable and Parallel Corpora (ECPC)**<sup>25</sup>, **Segurtur**

El término corpus comparable se emplea también para referirse a un corpus que consiste en diferentes variedades de un idioma según los grupos (EAGLES 1996a), como por ejemplo, el International Corpus of English (ICE)<sup>26</sup> que recopila inglés tanto oral como escrito de 13 países diferentes, o esTenTen<sup>27</sup> que consiste en un subcorpus de español de península y otro subcorpus de español americano, aplicando el mismo marco de diseño en ambos casos. El parámetro aplicado a esta comparación del idioma según regiones se denomina **núcleo o periferia** (Atkins et al., 1991, p.13).

- **Corpus Oral o escrito:** para algunos académicos un corpus oral es un corpus en el que se recogen conversaciones informales y espontáneas que han tenido lugar sin la intervención de ningún medio de comunicación. En otros casos, el uso del término se amplía para referirse a cualquier tipo de lengua en la que los hablantes se comportan de forma oral, como por ejemplo, en los textos escritos para ser hablados.

(Ejemplos)

Michigan Corpus of Academic Spoken English (MICASE)<sup>28</sup>, Santa Barbara Corpus of Spoken American English<sup>29</sup>, **Natural Conversation Corpus by BTS**<sup>30</sup>.

Se observa que tanto en español como japonés se disponen de varios corpora monolingües de referencia bastante completos en línea, y hay algunos corpora especiales, especializados y con fines específicos. En cambio, en caso de corpus bilingüe o plurilingüe, español tiene muchos corpora dentro del marco de la UE y el japonés tiene algunos con inglés, pero en cuanto a la combinación de estos dos idiomas, no se encuentra ninguno. A su vez, esta revisión indica la necesidad de considerar diferentes parámetros para diseñar un corpus.

A continuación se contrastan las características de dos tipos de corpora bilingües, corpus paralelo y corpus comparable. Después de estudiar las ventajas y desventajas de cada tipo, se determina el corpus más pertinente que se va a desarrollar en este trabajo.

## 2.2. *Corpus comparable y corpus paralelo*

Un corpus paralelo es la combinación de un corpus de textos escritos en un idioma y su traducción. Para los traductores son ejemplos reales de los diferentes traductores y les inspiran para llevar a cabo su trabajo. Presenta alta precisión a la hora de alinear, por eso se ha hecho muy popular para Procesamiento de Lenguajes Naturales (PLN) y para los sistemas de traducción automática (Teubert, 1996, p.246). Sin embargo, el corpus paralelo tiene un problema importante: la traducción

20 <http://www.statmt.org/europarl/>

21 [http://sli.uvigo.es/CLUVI/index\\_en.html](http://sli.uvigo.es/CLUVI/index_en.html)

22 <http://webcorpusak.elhuyar.eus/>

23 [http://actres.unileon.es/?page\\_id=33&lang=en](http://actres.unileon.es/?page_id=33&lang=en)

24 <https://www.gaois.ie/crp/en/>

25 <http://www.ecpc.uji.es/EN/home.php?language=en>

26 <http://ice-corpora.net/ice/>

27 <https://www.sketchengine.co.uk/documentation/tenten-corpora/>

28 <https://quod.lib.umich.edu/cgi/c/corpus/corpus?c=micase;page=simple>

29 <http://www.linguistics.ucsb.edu/research/santa-barbara-corpus>

30 <http://www.tufs.ac.jp/ts/personal/usamiken/corpora.htm>

es nunca libre de “*translationese*” (McEnery y Xiao, 2008, p.4). Según estos autores, la influencia de la lengua original (LO) es suficientemente fuerte para que el texto meta (TM) tenga una forma lingüística diferente que su forma natural (McEnery y Xiao, 2002 citado en McEnery y Xiao, 2008, p.6). Teubert (1996) también indica que “*Working with translations means working with distorted mirror images of the source language in the medium of the target language*” (p.247).

En cambio, un corpus comparable consta de una combinación de dos subcorpus monolingües que utilizan el mismo marco de diseño (McEnery y Xiao, 2008, p.2). Por lo tanto, tanto el subcorpus en lengua A como el de lengua B consisten en los textos originales y ofrecen los ejemplos reales de las lenguas sin “contaminación”. La mayoría de corpora comparables se desarrollan para el estudio de traducción y está demostrado que el acceso a los corpora comparables mejora la capacidad de traducción más fluida, sobre todo en el dominio especializado (Zanettin, 1998, citado en McEnery y Xiao, 2008, p.7). Para la extracción léxico se necesita más cantidad de texto, pero puede aplicarse el patrón de concurrencia (*co-occurrence pattern*) (Rapp, 1995, citado en Bo, 2012, p.18) o heterogeneidad contextual (*context heterogeneity*) (Fung, 1995, citado en Bo, 2012, p.19).

Tanto el corpus paralelo como el comparable son de utilidad y pueden ser complementarios uno al otro (Teubert, 1996, p.252; McEnery y Xiao, 2008, p.9; Bo, 2012, p.19). Sin embargo, desde un punto de vista práctico, un corpus paralelo necesita textos ya traducidos, lo cual en el estado actual de los recursos disponibles en español y japonés es muy difícil de conseguir. En cambio, los textos especializados están cada vez más accesibles tanto en español como en japonés, por lo que se considera apropiado desarrollar un corpus comparable especializado aprovechando estos recursos. A su vez, el acceso directo a los textos especializados en un idioma original fortalecerá el conocimiento sobre el lenguaje especializado de los traductores y así servirá para madurar la traducción de esta combinación lingüística.

### ***2.3. La calidad de un corpus comparable y su diseño***

A pesar de su utilidad, la definición de corpus comparable es todavía bastante ambigua y distintas corpora comparables pueden resultar muy diferentes (Bo, 2012, p. 5). Según Bo (2012), la noción de la calidad de corpus comparable, expresada como “comparabilidad”, tiene diferentes niveles. Por ejemplo, las temáticas que se tratan en los textos tienen que estar alineadas para desarrollar el estudio estadístico con la finalidad de descubrir la temática en un corpus (*topic model*) mientras que no es necesario para extraer léxico. Desde el punto de vista del desarrollo de PLN, Bo (2012) ubica el corpus paralelo como un tipo especial de corpus comparable. Indica que la experiencia en el uso de corpus bilingüe nos lleva a intuir que el grado de la comparabilidad degrada de un corpus paralelo (máxima comparabilidad) hacia un corpus no paralelo con la temática diferente (mínima comparabilidad) (p. 5). Basado en esta idea, la autora desarrolla unas medidas para cuantificar la comparabilidad que consiste en el criterio léxico basado en un diccionario y el criterio de desambiguación de sentido basado en el contexto. En cambio, según Goeuriot, Morin & Daille (2009) hay dos niveles de comparabilidad; el primer nivel es el entorno comunicativo como el dominio, la temática y los participantes, y el segundo nivel es el género (p.56). El objetivo de su investigación es desarrollar un algoritmo para automatizar la distinción de los géneros de texto, y menciona que la calidad del corpus depende mucho de los primeros pasos de la creación de mismo, la especificación del corpus y la selección de los textos (p.61).

Ahora con el fin de diseñar una metodología, se contrastan tres proyectos de creación de corpus comparable. El primero es el proyecto dentro de la UE, el European Comparable and Parallel Corpora (ECPC). Este proyecto partió de una hipótesis de “el grado de unificación europea es tan estrecho que instrumentos comunitarios, como los discursos de Parlamento Europeo (PE) poseen ciertas dosis de autonomía macrolingüística y/o microlingüística respecto de instrumentos

equivalentes en los Estados Miembros” (Calzada Pérez, Marín Cucala y Martínez Martínez, 2006, p.349) y desarrolló los corpora con el objetivo de alinear discursos (en caso del corpus paralelo) y generar concordancias y otro tipo de información similar entre los discursos del PE de la Cámara Baja del Parlamento británico, español e irlandés (en caso del corpus comparable) (ECPC). Los resultados se aplicaron en la enseñanza tanto de traducción como de la segunda lengua para los traductores e intérpretes.

El segundo es el corpus comparable de inglés y español, “Segurtur” (Seghiri, 2010). Su desarrollo tuvo el objeto de ser un recurso para estudios lingüísticos y traductológicos sobre la contratación de los seguros turísticos. Los componentes fueron las pólizas de seguro y éstos se seleccionaron por la vía “búsqueda institucional” y por la “búsqueda por palabras claves mediante buscadores generales”. La búsqueda institucional tuvo el mérito de encontrar gran cantidad de texto y textos con alta calidad y grado de fiabilidad, y se evaluó como una de las medidas más fructíferas (Seghiri, 2010). En cambio, la otra vía se aplicó de manera complementaria porque el resultado solía incluir muchos “ruidos”. El parámetro de evaluación que se aplicó al resultado fue su tamaño cuantitativo para asegurar su representatividad lingüística de este campo de especialidad.

El último trabajo es de Deléger (2008) que se trata de un caso de corpus comparable particular. Tuvo el objetivo de crear un corpus comparable del mismo idioma (francés) donde se pudiera ejercer una paráfrasis de textos especializados y populares de la misma temática médica (adicción a la nicotina). Aparte, estableció los objetivos específicos de investigar: 1. cómo recopilar textos más relevantes, 2. cómo identificar fragmentos que tengan información comparable y 3. qué tipo de paráfrasis podría conseguirse entre estos dos tipos de texto diferentes. Para la colección de textos, se acudió a un portal de salud cuyo buscador dirigió a los textos fiables filtrados según las palabras claves preparadas por el investigador y según el tipo de lector clasificado por el buscador. Los textos recopilados se alinearon manualmente según temas parecidos que posiblemente contuvieran paráfrasis. El corpus resultante se evaluó por su calidad (p. ej. si las paráfrasis detectadas fueron realmente correctas), por su tamaño (p. ej. las paráfrasis detectados fueron suficientes), y por la coherencia con el hipótesis de las características lingüísticas de los dos tipos de texto desarrollado por el mismo investigador. Aunque el corpus desarrollado es distinto que se desarrolla en este trabajo (corpus monolingüe con los diferentes tipos de texto), la descripción de la metodología detallada para la búsqueda de texto y la evaluación es ilustradora.

Cada corpus comparable está diseñado con diferentes objetivos (didáctica de la traducción, estudio lingüístico, estudio de la paráfrasis), criterios de selección de texto (por el escritor, por el tipo de texto y por la temática) y criterios de evaluación (la concordancia, la representatividad, el tamaño y la calidad lingüística). Esto indica la importancia de establecer objetivo claro y la noción de comparabilidad para desarrollar propio corpus comparable.

### **3. Marco teórico**

#### ***3.1. Criterios externos e internos para la selección de textos***

Como ya se ha observado en el epígrafe anterior, para la selección de textos se utilizan diferentes criterios. Entre los estudiosos también hay diferentes opiniones sobre qué criterios deben aplicarse para seleccionar textos. En los análisis lingüísticos se considera importante la distinción de los criterios externos e internos (Atkins et al., 1991, p.8). En caso de Sinclair (2005) subraya la importancia de emplear sólo los criterios externos y propone aplicar, como criterios básicos, la modalidad del texto (producción original escrita u oral), el tipo de texto (un libro, una revista, una carta etc.), el ámbito (académico o popular), el idioma, el lugar de producción, el tiempo de publicación. A estos criterios pueden sumarse los participantes (emisor y receptor), las

circunstancias, el entorno social y la función comunicativa (EAGLES, 1996b). Sinclair (2005) indica que si se aplican los criterios internos, se buscan textos que reflejen las características ya conocidas, por lo tanto, la selección estaría sesgada. Igualmente Atkins et al. (1991, p.8) considera problemático utilizar sólo los criterios internos porque el corpus no aporta información sobre la relación entre la lengua y su contexto, por lo tanto, pierde el valor crítico que ofrece el corpus, esto es, “el uso real de la lengua”. Asimismo, buscar textos por la temática supone cierta dificultad porque los textos suelen tratar temas transversales y, además, la percepción de la temática puede resultar subjetiva o incluso sociolingüística.

Sin embargo, en este punto se necesitaría considerar si son suficientes estos criterios externos para seleccionar textos especializados que se quieren compilar en este trabajo. Si se necesita recopilar los textos especializados, los criterios de selección de texto tienen que satisfacer ambas necesidades tanto de corpus como de texto especializado.

Hoy en día no hay una definición definitiva y unificada del concepto “lenguaje especializado” debido a que la delimitación entre la lengua natural y la lengua especializada es complicada. El lenguaje especializado y el lenguaje natural tienen la misma estructura en común y una palabra general se convierte en un término en ciertos ámbitos (Imamura, 2014). Según Cabré (1994), hay tres tipos representativos de definición del lenguaje especializado:

1. *La que defensa que els llenguatges especialitzats són codis de caràcter lingüístic, diferenciats del llenguatge general, que consten de regles i unitats específiques.*
2. *Els llenguatges d'especialitat són cada un d'ells simples variants del llenguatge general*
3. *Els llenguatges especialitzats com a subconjunts, fonamentalment pragmàtics, del llenguatge entès en sentit global (p.122)*

La autora parte de la tercera definición e indica las tres características comunes del lenguaje especializado centradas en el aspecto pragmático. La primera es la temática que no forma parte de los conocimientos generales. La segunda son los usuarios; el emisor siempre tiene conocimiento específico, aunque el del receptor varía desde el nivel estudiantil hasta el nivel profesional. La tercera es la situación de comunicación, que normalmente es formal y está dirigida por profesionales o científicos.

En cambio, Ciapuscio (2003) distingue los niveles funcional, situacional, de contenido semántico y formal-gramatical. Lo que se destaca en la posición de esta autora es que indica que la integración de los rasgos y valores de los distintos niveles dará como resultado una caracterización tipológica del texto:

Las unidades y relaciones en el nivel de la microestructura (léxico y gramática) están condicionadas por factores de tipo textual superior (básicamente, factores funcional-comunicativos y temáticos) y, por otro lado, estos factores de orden superior son parcialmente aseguibles (puede alcanzarse) y sistematizables a partir de los rasgos del nivel microestructural.

Esta autora investiga los diferentes niveles de especialización entre un resumen de un artículo científico, una crítica de un artículo científico, una entrevista y un periódico de un tema especializado, y clasifica las características de cada texto como lo siguiente:

		Texto Especializado		Texto No Especializado	
		Abstracto de un texto científico	Crítica	Entrevista	Periódico
Nivel Funcional	Expresar / Contactar / Informar / Dirigir	Informar (dominante) y Dirigir		Informar (dominante) y Dirigir	
	Jerarquía Funcional				
Nivel Situacional (Factores Ambientales Directos)	Comunicación Interna / Externa	Interna		Externa	
	Número de Interlocutores	Grupo Pequeño		Grupo Numeroso	
	Interlocutores	Especialistas (Relación Simétrica)		Interlocutor-Especialista/Lego Lector-Lego (Relación Asimétrica)	
	Parámetro	Publicación		Revista	Noticia
Nivel Contenido-Semántico	Tema del Texto	Contenido Denso			
	Actitudes Temáticas	Positiva, Certeza, Expositiva, Narrativa, Argumentativa		Expositiva, Narrativa	Expositiva, Narrativa
	Perspectiva sobre el Tema	Teórica		Divulgativa	
	Primaria/Derivativa	Primaria	Derivativa	Derivativa	Derivativa
	Partes Textuales	Estructura del artículo científico		Esquema Establecida	Parte Canónica
Nivel Formal	Máximas de Formulación de la Clase Textual	Economía, Desagentivación, Objetividad	Gramaticales Propios del Discurso	Personalización, Metáfora, Comparaciones	Foto, Distintas Tipografía
	Aspectos Gramaticales	3a Persona	1a y 3a Persona	1a Persona Singular	3a Persona
	Recursos Sintácticos y Léxicos	Pasiva, Nominalización	Verbos de Reporte a Opinión	Historia Personal	Discurso Referido
		Abundante Terminología sin Tratamiento	Alta Densidad Terminológica	Pocas Unidades Léxica Especializada con mucho tratamiento	Evitar Terminología
Nivel de Especialización		Muy alto (I)	Muy alto (I <sup>o</sup> )	Bajo (II)	Muy bajo (III)

Tabla 1. Clasificación de los niveles de especialización según el tipo de texto basada en Ciapuscio (2003).

Se considera muy razonable su indicación. Es cierto que hasta cierto punto las condiciones pragmáticas condicionan la forma lingüística de un texto; sin embargo, si un texto no cumple algunos rasgos lingüísticos que se supone que tiene que cumplir, tampoco va a percibirse tal como se supone por los aspectos pragmáticos. Cierta cumplimentación de condiciones tanto de la parte pragmática como de la parte lingüística es imprescindible para que un texto se perciba como un tipo de texto.

Teniendo en cuenta las condiciones de los niveles diferentes que deben satisfacer los textos especializados, seleccionar textos sólo por los criterios externos se considera discutible. Para distinguir los niveles tanto contenido-semántico como formal, se necesitan los criterios internos. Además, si se aplican solo los criterios externos, puede que dividamos textos que compartan similitudes lingüísticas y deban estar incluidos en el mismo corpus (EAGLES, 1996b). Por ejemplo, hay casos que no es razonable separar un texto didáctico publicado en línea del otro publicado en un libro. Por consiguiente, este trabajo toma la posición de aplicar ambos criterios de manera equilibrada. Los criterios externos facilitarían las primeras filtraciones de textos porque si se empieza por los criterios internos, se necesita leer los diferentes tipos de materiales y clasificarlos por el tema o por el lenguaje según el juicio de la persona. En cambio, los criterios externos delimitan bastantes características de texto de manera bastante clara y aseguran la imparcialidad de la selección del texto. Sin embargo, luego sería necesario aplicar los criterios internos que caracterizan el texto, como el lenguaje o la temática, para asegurar que los textos sean del mismo tipo y comparables uno al otro.

### ***3.2. Requisitos básicos de un corpus: cantidad, simplicidad y documentación***

Aparte de los criterios de selección de textos arriba mencionados, existen tres requisitos que deben ser cumplidos para que la colección de textos se considere como un corpus (EAGLES, 1996a). Son la cantidad, la simplicidad y la documentación. A continuación se revisan estos criterios brevemente.

### 3.2.1. Cantidad

Como se observa en su definición, el objetivo principal de corpus es el estudio de una(s) lengua(s). Para que un corpus pueda aportar datos útiles se necesita una gran cantidad de texto para buscar, recuperar, clasificar y calcular los contenidos con una velocidad inmensa. Es por eso que un corpus que no cuenta con los datos suficientes se considera que aporta sólo “the trivial level of studying alphabetic-character sequences” (Leech, 1991). Este mismo autor indica la conexión muy cercana entre “*corpus linguistics*” y “*quantitative linguistics*”. EAGLES (1996b) también apoya esta posición y añade que el tamaño de corpus sigue creciendo muy rápidamente incluso con la aparición de los *monitor corpora*. Pérez (2002), por su lado, hace mención a la polémica de “calidad vs. cantidad”. Apunta que ninguna postura debe ser llevada a extremos, aunque reconoce la practicidad y la ventaja de recolectar la mayor cantidad posible de los textos:

La postura contraria (de dar mayor importancia a la calidad) (expresada coloquialmente con la divisa “*more data is better data*” o “*there is no text like more text*”) puede ser [...] mucho más realista y ajustada a nuestro grado de conocimiento del uso de la lengua y a la realidad tecnológica y económica en la que vivimos [...] cuanto mayor sea el corpus, más posibilidades tendremos de que nos ofrezca información sobre un espectro más amplio de fenómenos lingüísticos [...] por ejemplo, la lexicografía: un diccionario como el OED contiene 250.000 entradas y un diccionario medio para estudiantes una media de 50.000 palabras, por lo que acumular evidencias lingüísticas [...] sobre un número tan elevado de entradas requiere, sin duda, que el corpus sea, por decirlo de alguna forma, cuanto más grande mejor.

Otro punto importante indicado por EAGLES (1996b) y Atkins et al. (1991, p.7) es la correlación de la cantidad de textos y la accesibilidad / el número de los usuarios de los materiales. Cuando se trata de textos populares como periódicos, revistas o radio (en el caso de un corpus oral), se requiere más cantidad para acceder a la mayor variedad posible. En cambio, si son materiales que circulan en una comunidad pequeña, la variedad de lenguaje es mucho más limitada y se requiere menos textos. Según Sinclair (2005) la mitad de las palabras en los textos se utiliza solo una vez, y un cuarto de palabras, dos veces. Estudiar palabras que aparecen solo una o dos veces tiene muy poco sentido en la lingüística. El autor compara el LOB (corpus de referencia) y English of Computing Science (HK, corpus especializado) que contienen la misma cantidad de palabras (un millón) y concluye que la cantidad de textos necesarios para el estudio lingüístico de textos especializados puede ser mucho menor que un corpus general (2005).

	LOB	HK	%
Number of different word-forms (types)	69990	27210	39%
Number that occur once only	36796	11430	31%
Number that occur twice only	9890	3837	39%
Twenty times or more	4750	3811	80%
200 times or more	471	687	(69%)

Tabla 2. Comparación de las frecuencias léxicas en un corpus general y un corpus especializado extraído de Sinclair (2005).

Se observa que en el HK la variedad de las palabras es mucho menor que en el LOB, sin embargo, un 14% de las palabras se repiten más de 20 veces (en caso de LOB sólo un 6 %) y un 2% de las palabras, más de 200 veces (en caso de LOB, un 0,6 %). Esto indica que la variación terminológica es menor y los mismos términos se emplean con mucha frecuencia a lo largo de todo el artículo.

Aparte, hay discusión sobre si los textos que forman un corpus deben emplearse enteros o parciales. Esta cuestión se estudia desde dos puntos de vista: la calidad del corpus y los derechos de la propiedad intelectual de los autores. Por un lado, como indica Pérez (2002), seleccionar u omitir un texto puede repercutir en la información conceptual del mismo, por lo que, los textos individuales deben ser textos completos. Sinclair (2005) apoya esta posición y recomienda incluir los textos enteros, aunque también indica que si los textos grandes en un corpus pequeño afectan a la representatividad del corpus, entonces hay que seleccionar una parte del texto según el “*best guess*” del creador de corpus. Para EAGLES (1996a), la condición predeterminada debe ser textos enteros y considera un corpus de fragmentos de texto como un corpus especial .

En realidad, se observan muchos corpora que no siguen a esta instrucción. Por ejemplo, los corpora clásicos como el ICE o LOB consisten en los fragmentos de textos de aproximadamente 2.000 palabras. El diseño puede que esté motivado por la disponibilidad tecnológica de la época (estos proyectos empezaron en 1990 y 1978 respectivamente), pero por sus cuidadosos diseños están considerados como corpora de referencia representativos del inglés estándar (Leech, 1991). El corpus KOTONOHA también selecciona partes de los textos de manera aleatoria (*random sampling*), por un lado, según el número de caracteres (1.000 caracteres), y por otro lado, según párrafo o capítulo. Se toma esta medida con el objeto de recolectar los datos más imparciales y descubrir la mayor variedad y la representatividad posible.

Por otro lado, utilizar el texto entero de un tercero levanta la cuestión de los derechos de la propiedad intelectual de los autores. Para utilizar un texto entero normalmente se necesita permiso del autor, lo cual implica una dificultad tanto por el tiempo como el labor requeridos para obtenerlo (Atkins et al, 1991, p. 4; Sinclair, 2005; Rodríguez Inés, 2008, p.31). En cambio, cuando se utilizan fragmentos de obras de otras ajenas de naturaleza escrita ya divulgadas con el objeto de utilizarlo en las investigaciones académicas y no lucrativas, se permite su uso sin autorización siempre y cuando se proporcione la referencia bibliográfica (Da Cunha, Torres Moreno, Sierra, Cabrera Diego, Castro Rolón & Rolland Bartilotti, 2011, p.700; Galinski y Wright 1997 citado en Pérez, 2002). El tema de los derechos de la propiedad intelectual debe tomarse con seriedad al desarrollar un corpus.

### 3.2.2. Simplicidad

Los textos componentes de un corpus deben ser guardados en el formato texto simple (.txt), texto enriquecido (.rtf) o el documento (.doc) para que después puedan ser tratados por editores de texto, codificados y anotados. Sin embargo, la información añadida (la anotación o la codificación) a los textos originales deben ser totalmente separables e identificables.

### 3.2.3. Documentación

La referencia bibliográfica de los textos componentes del corpus debe ser informada por la cuestión de los derechos de la propiedad intelectual como antes mencionado (3.2.1.). También la referencia debe ser guardada de manera separada de los textos (Leech, 1991; EAGLES, 1996a).

## 4. Metodología

### 4.1. El corpus

La metodología aplicada en este trabajo está inspirada por los trabajos de Deléger (2008) y de Seghiri (2010) con algunas modificaciones. El proceso consiste en: 1. Diseño de corpus, 2. Búsqueda y selección de los textos, 3. Normalización de los textos, 4. Almacenamiento y 5. Registro de la fuente de los textos componentes. A continuación se detallan las tareas realizadas en

cada proceso.

Fase	Acción	Detalle
1	Diseño de corpus	Establecer los criterios para conseguir el corpus deseado. Los criterios pueden ser: tipo de texto, lugar y año de publicación, la temática etc.
2	Búsqueda y selección de los textos	Seleccionar los textos en base a los criterios establecidos.
3	Normalización	Convertir el formato de texto en texto simple (.txt) para que sean manejables en cualquier editor de texto y para que sean tratables en el proceso posterior, como la concordancia o el etiquetado.
4	Almacenamiento	Guardar los textos correctamente identificados y relacionados.
5	Registro de fuentes de los textos	Registrar los datos necesarios para que sean identificables.

Tabla 3. Lista de tareas para el desarrollo del corpus.

#### 4.1.1. Diseño de corpus

Antes de compilar los textos para el corpus, es imprescindible establecer la noción de la comparabilidad que quiere obtenerse. Los objetivos del corpus en esta investigación pueden resumirse en dos puntos: servir como una herramienta práctica para la traducción y fortalecer el lenguaje especializado de los traductores. Para el primer objetivo, se considera relevante que los textos recopilados tengan léxicos y términos equivalentes. Aparte, se pretende ofrecer unos fragmentos que tengan la alta similitud de información en ambos idiomas para que el traductor pueda aprovechar tanto la semántica como la sintaxis. Para ello, la coincidencia del contexto sería importante. Para el segundo objetivo, este trabajo pretende extraer textos de lenguaje de especialización más alto que corresponde al I (muy alto) de la tabla 1.

Para evaluar estos valores de la comparabilidad, se establecen los siguiente parámetros:

- lenguaje (se emplea el lenguaje del mismo nivel de especialización)
- término (se encuentran los términos equivalentes)
- fragmento (se encuentran los fragmentos con alta similitud)

Para conseguir un corpus que satisfaga esta comparabilidad, se establecen siguiente criterios del diseño del corpus:

Tipo	académico
Emisor	especialista del ámbito
Receptor	especialista-semilego
Idioma	español (España) y japonés (Japón)
Modo	escrito (original: escrito, interpretación: escrito)
Medio	texto publicado (en el formato papel o en el formato electrónico)
Texto completo/fragmento	fragmento (por capítulos o por epígrafes)
Año de publicación	a partir del 2010
Temática	innovación tecnológica

Tabla 4. Criterios establecidos para la selección de texto.

De los diferentes ámbitos especializados, este trabajo enfoca en la innovación tecnológica porque es uno de los sectores donde se necesita una comunicación cada vez más dinámica. Es un ámbito que se ha convertido en la base de nuestra sociedad y la industria es muy activa a nivel mundial. Por lo

tanto, los traductores tienen más ocasiones de trabajar en este ámbito. La cantidad mínima de los textos en cada idioma se establece como cinco para garantizar la mínima variedad terminológica e idiolecto. La selección de textos se realiza por capítulos o epígrafes para conseguir mayor densidad de comparabilidad. Es preferible que el tamaño de textos sea igual, sin embargo, en esta investigación se decide no establecer un tamaño de texto concreto para evitar la recopilación de textos poco comparables.

Antes de seleccionar los textos, se recopilan las características del lenguaje especializado en cada idioma y se resumen en una tabla (anexo 1).

#### 4.1.2. Selección de los textos

Para asegurarse de que los textos cumplieran todos los criterios diseñados, se considera adecuado tratar las tesis doctorales publicadas en línea. Por lo tanto, se acude a los buscadores de las tesis doctorales, Tesis Doctorals en Xarxa (TDX)<sup>31</sup> para los textos en español y CiNii Dissertations<sup>32</sup> para los textos en japonés. TDX es un repositorio de aproximadamente 30.000 tesis publicadas por 18 universidades españolas. CiNii Dissertations también es un buscador de tesis doctorales publicadas en Japón y contiene información de 600.000 tesis y tiene acceso a los 130.000 textos.

Llegados a este punto se necesita plantear una metodología propia para seleccionar textos porque ninguna metodología empleada en los trabajos anteriores es válido: por un lado, en el trabajo de Delégar primero se elabora una lista de palabras claves para buscar textos en los buscadores institucionales. Sin embargo, de esta manera los resultados tanto de TDX como de Cinii incluyen gran cantidad de textos no pertinentes, o en caso contrario, no extraen ningún documento resultante. Por otro lado, los textos buscados en el corpus de Seghiri son las pólizas de seguro turístico, que es un tipo de texto bastante estandarizado tanto por su estructura como contenido. Por lo tanto, es diferente que buscar tesis doctorales donde la temática tratada y su estructura son diferentes según cada texto. Por consiguiente, se ajusta la búsqueda a la siguiente manera: primero se filtran los textos con la palabra clave, “人工知能 (inteligencia artificial)” en Cinii. Como las palabras claves deben aparecer en el resumen (Da Cunha, 2015, p.88), desde los textos seleccionados se analizan los resúmenes de cada texto en el orden del año de publicación descendiente y se extrae una palabra clave más para cada texto<sup>33</sup>. Estas dos palabras claves traducidas en español se aplican en la búsqueda del texto comparable en TDX para filtrar textos adecuados. Desde los textos resultantes, se revisan los resúmenes en el orden del año de publicación descendiente y se selecciona un texto comparable, que realmente contenga las palabras claves y cuya temática tenga algún solape con la del texto japonés. Cuando no se encuentra ningún texto comparable en español, se descarta el texto japonés y se empieza una nueva búsqueda. Se repite este proceso hasta que se consigan 5 pares de textos comparables.

Una vez seleccionados los textos, se contrastan los índices entre español y japonés para buscar capítulos o epígrafes más comparables. Según el texto, algunos permiten el acceso solo al resumen y algunos epígrafes contienen muy pocos textos, por lo tanto, la cantidad de texto extraído es diferente en cada caso. De todas formas, se seleccionan solo las partes más interesantes donde se encuentran los términos en común en ambos idiomas.

#### 4.1.3. Normalización

La mayoría de los textos están publicados en PDF (.pdf) o en HTML (.html). En esta fase, estos formatos deben convertirse al texto simple (.txt) para los tratamientos posteriores, como por ejemplo, codificación o anotación.

31 <http://www.tdx.cat/>

32 <https://ci.nii.ac.jp/d/?l=en>

33 Se seleccionan primero los textos japoneses porque suelen haber más restricciones en el acceso a los documentos en Japón.

En este trabajo, primero el formato se convierte en el Texto OpenDocument (.odt) para realizar el tratamiento preliminar como contar palabras o eliminar los gráficos, las formulas matemáticas y los símbolos ( $p$ ,  $d_{pp}$ ,  $\beta$  etc.) porque desde el punto de vista de traducción no son relevantes y, además, no se soportan en el texto simple. Para convertir el formato en el Text OpenDocument, se acude a los programas de conversión de pdf como iLovePDF<sup>34</sup> y Wondershare<sup>35</sup>.

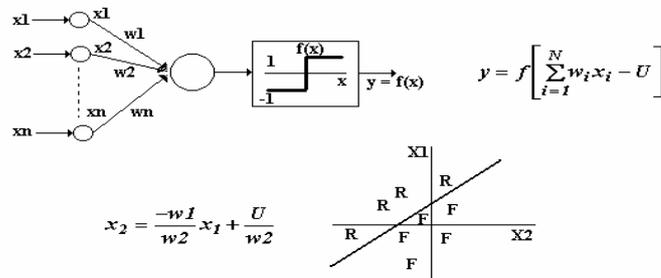


Figura 1. Ejemplos de los gráficos eliminados extraídos de Peguero Chamizo (2003) del texto 5E.

Después de ser limpiados, todos los textos se guardan en el formato texto simple (.txt).

#### 4.1.4. Almacenamiento

En esta fase los documentos descargados y convertidos se guardan correctamente identificados.

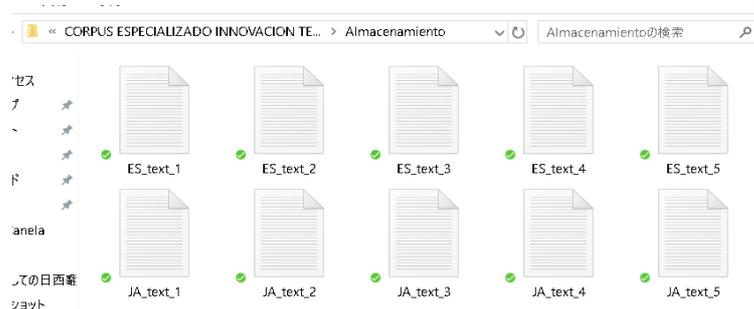


Figura 2. Almacenamiento de los documentos para el corpus.

#### 4.1.5. Registro de la fuente de los textos componentes

Este proceso no se especifica en los trabajos de Deléger ni de Seghiri, sin embargo, se considera apropiado apuntarlo como una fase por su importancia. Como se ha indicado en un epígrafe anterior (3.2.1.), es necesario documentar la fuente de los textos para poder utilizarlos en el corpus creado, respetando los derechos de los autores. Por lo tanto, después de almacenar los textos componentes, se registra la información del autor, la institución a la que pertenece y el año de la publicación de todos los textos (anexo 2).

### 4.2. El glosario

#### 4.2.1. Extracción de los términos

Aprovechando el corpus creado, se extraen términos y crea un glosario bilingüe especializado. Para la extracción de los términos se disponen de varias herramientas en las instituciones académicas,

34 <https://www.ilovepdf.com/ja>

35 <https://www.wondershare.es/>

por ejemplo, Terminus de la Universitat Pompeu Fabra (UPF)<sup>36</sup>, Tes-wizard de la Universitat Oberta de Catalunya<sup>37</sup> o Gensen Web de la Universidad de Tokio<sup>38</sup>. Sin embargo, en este trabajo este proceso se realiza manualmente porque en cada herramienta se observan fallos en la extracción. Por ejemplo, en el Gensen se extrae fragmentos como “clínica humana según” o “リモートかつメンテナンスフリー remoto y libre de mantenimiento” como términos. Desde cada texto se alistan los términos (anexo 3).

#### 4.2.2. Búsqueda de equivalencia

Cada lista de términos se contrasta con la lista de su texto comparable y se extraen los que tienen su equivalencia en la otra lista. Después de revisar cada texto comparable, se examinan todos los términos con equivalencia para detectar y eliminar los términos duplicados. Al final se crea una lista de los términos que van a ingresarse en el glosario bilingüe (anexo 4).

#### 4.2.3. Ingreso de datos

A día de hoy, se disponen de varias herramientas en línea que tienen la función de crear un glosario como, por ejemplo, SDL Trados<sup>39</sup>, Memosource<sup>40</sup>, OmegaT<sup>41</sup>. En este trabajo se utiliza OmegaT porque el sistema ofrece una interfaz sencilla e intuitiva, y además, es una herramienta muy familiar entre los traductores. Los términos de la fase anterior se ingresan en el sistema y, como resultado, se crea un glosario en el formato de texto simple (.txt).

### 5. Resultado y análisis

#### 5.1. Resultado cuantitativo

Como resultado de la recopilación de los 5 textos en cada idioma, se crea un corpus de 18.929 caracteres en japonés y 9.093 palabras en español. Se recopilan 367 términos en japonés y 314 términos en español, de los cuales se registran 102 términos en japonés y 89 términos en español traducibles de uno al otro en el glosario.

Texto	Términos en japonés	Términos en español	Términos con equivalencia
1	86	77	JA 24 / ES 20
2	45	32	JA 8 / ES 8
3	40	56	JA 8 / ES 8
4	51	49	JA 10 / ES 6
5	150	112	JA 54 / ES 50
Total	367	314	JA 102 / ES 89

Tabla 5. Los términos extraídos de los textos recopilados.

Los números totales de los términos extraídos no son la suma de los términos de cada texto porque entre los diferentes textos habían algunos términos solapados. Tampoco los números de los términos en común entre español y japonés son los mismos porque algunos términos tienen varias formas diferentes. Por ejemplo, para el término “conjunto difuso”, en japonés se detectan dos términos

36 <http://terminus.iula.upf.edu/cgi-bin/terminus2.0/terminus.pl>

37 <http://traduccio.blogs.uoc.edu/2012/04/13/52/>

38 [http://gensen.dl.itc.u-tokyo.ac.jp/gensenweb\\_eng.html](http://gensen.dl.itc.u-tokyo.ac.jp/gensenweb_eng.html)

39 <http://www.sdl.com/es/software-and-services/translation-software/sdl-trados-studio/>

40 <https://www.memosource.com/>

41 <http://omegat.org/>

“Fuzzy Set” y “ファジィ集合”.

4	学習	aprendizaje	4
5	unsupervised learning	aprendizaje no supervisado	5
6	教師無し学習		
7	supervised learning	aprendizaje supervisado	6
8	教師あり学習		
9	Back Propagation法		7
10	BP法	Backpropagation	
11	バックプロパゲーション法	Propagación hacia atrás del error	8
12	逆誤差伝播則		
13	層	capa	9
14	入力層	capa de entrada	10

Figura 3. Un fragmento de términos almacenados en el glosario.

En el caso del japonés, un 27 % de la variedad terminológica ha encontrado sus traducciones correspondientes en español, y en caso de español, un 28 %. Si se tiene en cuenta que el alto grado de repetición de los mismos términos es una de las características de los textos especializados (Sinclair, 2005), el total de las coincidencias terminológicas debe ser más alta.

## 5.2. Resultado cualitativo

Entre los textos recopilado, se observan algunos fragmentos que tratan exactamente el mismo tema. Por ejemplo, en los textos comparables 5, se encuentran los siguientes fragmentos:

(ES5-1) “Una Red Neuronal es, básicamente, una estructura en la que se combinan un conjunto de elementos de cálculo que, por similitud con el modelo biológico en el que se basan, se denominan neuronas.”

(JA5-1) “Una Red Neuronal es un conjunto complejo en el que se reúnen varios elementos de proceso de información que se basan el modelo biológico, neuronas. Y ...”

(“ニューラルネットワークとは生物の脳神経系すなわちニューロンを模擬した単純な情報処理素子を多数個集めて複雑に結合させ...”)

o

(ES5-2) “No supervisado: la red no recibe influencia externa para ajustar sus pesos. Se dice que estas redes tienen capacidad autoorganizativa.”

(JA5-2) “En caso de aprendizaje no supervisado (la red) no recibe el señal para ajustar los pesos desde externo, por lo que necesita tener incorporada la capacidad autoorganizativa.”

(“教師なし学習の場合は、理想的な出力は外部から与えられないので自分自身の評価基準を内蔵しておくことが必要となる。”)

Estas frases no son exactamente iguales. Sin embargo, la información tiene una similitud muy alta. Si se necesita traducir estos fragmentos, se puede utilizar los fragmentos equivalentes de su texto comparable con muy pocas modificaciones.

En cuanto al lenguaje y la densidad terminológica, tanto en español como en japonés se cumple el hipótesis del lenguaje especializado alistado en el anexo 1 (actitud positiva y expositiva, la formulación con economía, el uso de tercera persona singular y pasiva, la finalización de frase con “DEARU” etc.) y se emplean abundantes términos. Por estas observaciones, los textos compilados se consideran que satisfacen una buena calidad.

## 6. Conclusiones

Un corpus es un recurso muy potente tanto en la lingüística como en la traductológica. Dentro de los diferentes corpora, especialmente corpus comparable es un recurso que se desarrolla para diferentes objetivos en diferentes metodologías con diferentes parámetros de evaluación. En la metodología aplicada en esta investigación, la búsqueda y la recopilación de textos han conllevado cierta dificultad. Sin embargo, basado en el análisis cuantitativo y cualitativo con los parámetros de evaluación diseñados, los textos seleccionados presentan una alta comparabilidad tanto en los términos como en los fragmentos más grandes, lo cual indica su validez como un recurso para la traducción especializada. Con el fin de ampliar su capacidad, se continuará la recopilación de textos. También sería interesante expandir la temática del corpus de cara al futuro.

## 7. Bibliografía

- Atkins, S., Clear, J., Ostler, N. (1991). Corpus Design Criteria. *Literary and Linguistic Computing*, 7(1), 1-16. DOI: 10.1093/lc/7.1.1
- Bo, L. (2012). Measuring and Improving Comparable Corpus Quality. Recuperado de <https://tel.archives-ouvertes.fr/tel-00997769/document>
- Cabré, M.T. (1994). *La Terminologia. La Teoria, els Mètodes, les Aplicacions*. Barcelona: Empúries. ISBN: 84-7596-363-3
- Calzada Pérez, M., Marín Cucala, N. & Martínez Martínez, J. M., (2006, septiembre). ECPC: European Parliamentary Comparable and Parallel Corpora / Corpus Comparables y Paralelos de Discursos Parlamentarios Europeos. *Procesamiento del Lenguaje Natural*, vol. 37, 349-350. ISSN: 1135-5948.
- Ciapuscio, G.E. (2003). *Textos Especializados y Terminología*. Barcelona: UPF. ISBN: 84-477-0855-1.
- Da Cunha, I. (2015). *El Trabajo de Fin de Grado y de Máster: Redacción, Defensa y Publicación*. Barcelona: Oberta UOC Publishing.
- Da Cunha, I., Torres Moreno, J. M., Sierra, G., Cabrera Diego, L. A., Castro Rolón, B. G. & Rolland Bartilotti, J. M. (2011). The RST Spanish Treebank on-line Interface. En G. Angelova (Ed.), *Trabajo presentado en Recent Advances in Natural Language Processing. International Conference 2011 (RANLP 2011)*, pp. 698-703. Hissar: ACL. ISBN: 978-1-62276-511-9
- Deléger, L., Zweigenbaum, P. (2008, febrero). Paraphrase Acquisition from Comparable Medical Corpora of Specialized and Lay Texts. *AMIA Annual Symposium Proceedings, 2008*, 146-150. PMID: PMC2656025
- EAGLES (1996a). *Preliminary Recommendations on Corpus Typology*. Retrived from <http://www.ilc.cnr.it/EAGLES96/corpusTyp/corpusTyp.html>
- EAGLES (1996b). *Preliminary Recommendations on Text Typology*. Retrived from <http://www.ilc.cnr.it/EAGLES96/textTyp/textTyp.html>
- ECPC (n.d.). En ECPC – European Comparable and Parallel Corpora. Recuperado de <http://www.ecpc.uji.es/ES/home.php?language=es>
- Fantinuoli, C. (2016, junio). Revisiting Corpus Creation and Analysis Tools for Translation Tasks. *Cadernos de Tradução*, Vol. 36, 62-87. ISSN 2175-7968.
- Goeriot L., Morin E. & Daille B. (2009, agosto). Compilation o Specialized Comparable Corpus in French and Japanese. *Proceedings of the 2nd Workshop on Building and Using Comparable*

*Corpora, ACL-IJCNLP 2009*, 55-63. DOI: 10.3115/1690339.1690353.

- Gómez de Enterría, J. (2010). El Lugar que Ocupan las Lenguas de Especialidad en la Enseñanza y Aprendizaje del Español como Lengua Extranjera. En V. Agustín e I. Martínez (Eds.), *El Español en Contextos Específicos: Enseñanza e Investigación*, pp. 41-64. Comillas: Fundación Comillas.
- Imamura, K. (2014, marzo). Identification of Field Specific Terms, Common Terms of Social Sciences, and Common Academic Terms in Fundamental Literature in Social Sciences. *専門日本語教育研究* 16 (0), 29-36.
- Leech, G. (1991). The State of the Art in Corpus Linguistics. En K. Aijmer K. y Altenberg B. (Eds.), *English Corpus Linguistics: Linguistic Studies in Honour of Jan Svartvik*, pp.8-29. Londres: Longman,
- McEnery A., Xiao Z. (2008) Parallel and Comparable Corpora: What are they up to? En G.M. Anderman y M. Rogers (Eds.), *Incorporating Corpora: Translation and the Linguist*. Clevedon: Buffalo. ISBN 978-1-85359-986-6
- Pérez Hernández, M. C. (2002). Explotación de los Córpora Textuales Informatizados para la Creación de Bases de Datos Terminológicas Basadas en el Conocimiento. *Estudios de Lingüística del Español (EliEs)*, vol. 18. ISSN: 11398736
- Rodríguez Inés, P. (2008). *Uso de Corpus Electrónicos en la Formación de Traductores (Inglés-Español-Inglés)*. UAB, Barcelona.
- Sakaue, T. (2015, diciembre). Compilation of a New Learner Corpus for Second Language Acquisition Research. *Bolletín of the Graduate School of Education Part II, Arts and Science Education*, Vol. 64, 139-146. ISSN: 1346-5554
- Seghiri, M. (2010). Metodología de Diseño y Compilación de un Corpus Representativo de Seguros Turísticos. En R. López-Campos Bodineau, C. Balbuena Torezano & M. Álvarez Jurado (Eds.), *Traducción y modernidad: Textos Científicos, Jurídicos, Económicos y Audiovisuales*, pp.59-70. Córdoba: UCO. ISBN: 9788499270432
- Sinclair, J. (2005). Corpus and Text - Basic Principles. En M. Wynne (Ed.), *Developing Linguistic Corpora: a Guide to Good Practice*, pp. 1-16. Oxford: Oxbow Books. ISSN 1463 5194 D
- Teubert, W. (1996, septiembre). Comparable or Parallel Corpora? *International Journal of Lexicography*, 9 (3), 238-264. DOI: 10.1093/ijl/9.3.238

## 8. Anexos

Anexo 1: Características de lenguaje especializado

Anexo 2: Registro de textos del corpus

Anexo 3: Términos extraídos

Anexo 4: Términos registrados en el glosario

## Anexo 1. Características de lenguaje especializado

Español	Japonés
- Tercera persona singular (primera persona plural)	- Omisión de sujeto
- Verbos en presente indicativo (verbos en pasado)	- Verbos en presente (conocimiento en común) Verbos en pasado (resultado de experimento)
- Verbos modales y perífrasis	
- Pasiva	- Pasiva
- Sintagma nominal / nominalización de los verbos	- Sintagma nominal intensa
- Frase larga con estructura interior compleja	- Frase corta con estructura interior compleja
- Oración declarativa con verbo copulativo en oración simple	- Oración declarativa y la finalización de frase con “DEARU” (である)
- Actitud positiva con certeza	
- Narrativa expositiva o aumentativa	- Abundante uso de términos (aumento del uso de la letra KATAKANA)
- Abundante uso de términos	
- Uso de ciertos verbos o conectores	- Uso de ciertos verbos o conectores

### Bibliografía

- Álvarez Jorge, P. (2008). *La Sintaxis del Lenguaje de los Textos Científicos. Los Tipos Oracionales y los Giros de Participio y Gerundio: Estudio de un Corpus Ruso-Español de Textos Médicos*. (Tesis de Máster inédita). Traducción, Interpretación y Estudios Interculturales. Universitat Autònoma de Barcelona. Barcelona.
- Cabré, M.T. (1994). *La Terminologia. La Teoria, els Mètodes, les Aplicacions*. Barcelona: Empúries. ISBN: 84-7596-363-3
- Ciapuscio, G.E. (2003). *Textos Especializados y Terminología*. Barcelona: UPF. ISBN: 84-477-0855-1.
- Halabis, W. (2014). *Español con Fines Académicos: Guía de Revisión 2.0 para la Escritura Académica*. (Tesis de Máster inédita). Enseñanza de Español como Lengua Extranjera.. Universidad de Cantabria. Cantabria.
- Murata, M. (2013). 社会科学系書籍における複合動詞の使用傾向：後項動詞を指標として. *日本語と日本語研究 No.41*, 67-95. Tokio. Keio University.
- NINJAL (1981). 専門語の諸問題. *国立国語研究所報告 68*. Tokio: 秀英出版.
- NINJAL (2013). テキストにおける語彙の分布と文章構造 成果報告書. *国立国語研究所共同研究報告 12-06*. ISSN 2185-0127
- Schifko, P. (2001) ¿Existen las Lenguas de Especialidad? En M. Bargalló, E. Forgas, C. Garriga, A. Rubio & J. Schnitzer. (Eds.) *Las Lenguas de Especialidad y su Didáctica*. Tarragona: Universitat Rovira i Virgili, pp. 21-29.

Anexo 2. Registro de los textos del corpus

	Palabras de la búsqueda	Título	Año de publicación	Institución	Autor	caracteres (JA)	palabras (ES)	
1	1J	人工知能メタデータ	Studies about Collaborative Annotation for Making Semantic Metadata (1.3.2. Semantic Web)	2008	The Graduate University for Advanced Studies, School of Multidisciplinary Science, Department of Informatics	Yuki Matsuoka	3.072	
	1E	Inteligencia Artificial Metadatos	Mecanismos semánticos orientados a la flexibilidad de los repositorios para objetos de aprendizaje (4.1. Web Semántica)	2008	Universidad de Alcalá, Departamento de Ciencias de la Computación	Soto Carrión, Jesús		1.294
2	2J	人工知能マルチモーダル	Multilayered Multimodal Latent Dirichlet Allocation for Probabilistic Knowledge Representation (Resumen)	2015	The University of Electro-Communications, Departamento de Ingeniería	Muhammad, Attamimi	2.846	
	2E	Inteligencia Artificial Multi modal	Incremental active learning learning of sensorimotor models in developmental robotics (Resumen en español) (ISBN 9788449052262)	2015	Universitat Autònoma de Barcelona, Departamento de Ciències de la Computació	Ribes Sanz, Arturo		605
3	3J	人工知能シミュレーション	Three Intelligent Systems for Supporting Advanced Technology – Multi-agent, Image Recognition, and Machine Learning- (Resumen 1- 2.1.4. )	2016	Kanazawa University, Graduate School of Science and Technology, Specialized in Science of electronic information	Masanichi Nitta	3.057	
	3E	Inteligencia Artificial Simulación	Simulación asistida por agentes para sistemas de fabricación inteligentes (3.5. La simulacion y los sistemas Multiagente, 3.5.1. El Proceso de Diseño de Simulación Basada en Agentes)	2009	Universidad Politécnica de Valencia, Departamento de Sistema informáticos y Computación	Ruíz Vega, Nancy		2.176
4	4J	人工知能ファジィ	Study on Reinforcement Learning System using Self-Organizing Fuzzy Neural Network (2.2 Fuzzy Neural Network, 2.2.1. Fuzzy Set 2.2.2. Composition of Fuzzy Set 2.2.3 Fussy Inference)	2014	Yamaguchi University, Graduate School of Science and Engineering	Takashi Kuremoto	1.18	
	4E	Inteligencia Artificial Regla difusa	Aprendizaje cooperativo para modelado lingüístico flexible basado en reglas difusas: interpretabilidad y precisión (1.1. Identificación de Sistemas, 1.2 Modelado de Sistemas con Sistemas Basados en Reglas Difusas)	2001	Universidad de Granada, Departamento de Ciencias de la Computación e Inteligencia Artificial	Casillas Barranquero, Jorge		1.606
5	5J	人工知能ニューラルネットワーク	Study about the application of neural network for the design of squirrel-cage rotor and his transient (3.2. Resumen of Neural Network, 3.3. Learning Algorithms in the Neural Network) *traducido por la autora	2005	Kumamoto University, Graduate School of Science and Technology	Masahiro Ikeda	8.774	
	5E	Inteligencia Artificial Red neural	Control de una planta de tratamiento de aguas residuales mediante redes neuronales, (Capítulo 2 Redes Neuronales, 2.1. Definición de Red Neuronal 2.3. Modelo práctico de red neuronal 2.3.1El Perceptrón, 2.3.2 Percepción Multicapa)	2003	Universidad de Extremadura, Escuela de Ingenierías Industriales, Departamento de Electrónica e Ingeniería Electromecánica	Peguro Chamizo, Juan Carlos		3.412

Total

Caracteres en japonés: 18.929

Palabras en español: 9.093

Anexo 3. Lista de términos extraídos  
(orden alfabético)

TEXTO 1

Términos en español			
1	agente autónomo	40	mecanismo
2	agente de la bolsa autónomo	41	metadato
3	agentes inteligentes	42	metalenguaje
4	aplicación	43	modelo conceptual
5	BCM02	44	modelo OIL
6	bolsa	45	módulo
7	capa Logic	46	módulo SWRLTAB
8	checksum	47	motor de inferencia
9	código de representación	48	nivel Trust
10	consulta	49	ontología
11	DAML	50	operación de autenticidad
12	definir	51	Operación de veracidad
13	Digital Signature	52	ordenador
14	documentos estructurados	53	OWL
15	elemento vertical	54	OWL-DL
16	escenario	55	página web
17	especificación	56	portal
18	esquema	57	proceso de inferencia
19	esquema básico	58	proveedor
20	esquema rdfrrs	59	razonador
21	estándar	60	regla
22	etiquetas de formato	61	regla lógica
23	fichero XML	62	reglas de inferencia
24	firma digital	63	representación
25	herramienta	64	restricción
26	herramienta de verificación del modelo	65	semántica
27	HTML	66	servidor
28	identificador	67	servidor público
29	implementación	68	sintaxis superficial
30	inferencia	69	sistema GIS
31	inferencia simple	70	software
32	lenguaje de descripción	71	SWRL
33	lenguaje de ontología	72	SWRL rule engine
34	lenguaje formal	73	vocabulario específico
35	lenguaje XML	74	web
36	lógica descripciones	75	web semántica
37	máquina	76	XML scheme
38	marco	77	XML Security
39	marco básico		

Términos en japonés			
1	HTML	11	<b>RDFS</b>
2	HTMLブラウザ	12	RDFサーバ
3	Java	13	<b>RDFスキーマ</b>
4	<b>Named Entity Recognition</b>	14	RDFトリプル
5	<b>object</b>	15	<b>Resource Description Framework</b>
6	<b>OWL</b>	16	<b>subject</b>
7	<b>predicate</b>	17	TAP
8	<b>RDF</b>	18	<b>Taxonomy Based Disambiguationアルゴリズム</b>
9	<b>RDF Scheme</b>	19	<b>TBDアルゴリズム</b>
10	<b>RDF Vocabulary Description Language 1.0</b>	20	Uniforme Resource Identifier

21	URI	54	ローカルマシン
22	URI参照	55	意味的メタデータ
23	<b>Web Ontology Language</b>	56	拡張
24	webコンテンツ	57	基本クラス
25	XML構文	58	基本プロパティ
26	Xpointer	59	機械
27	アノテーション	60	記述
28	アプローチ	61	記述言語
29	インスタンス	62	空白ノード
30	オントロジ	63	限定条件
31	クラスコンセプト	64	固有名詞
32	サブクラス	65	<b>固有名認識</b>
33	サーバ	66	語彙関係
34	スキーム	67	高機能化
35	セマンティック	68	自動アノテーションツール
36	セマンティックweb	69	自動マークアップ
37	セマンティックレポジトリ	70	識別子
38	ツール	71	<b>主語</b>
39	テキスト分析プラットフォーム	72	<b>述語</b>
40	データモデル	73	処理
41	データ記述モデル	74	書式
42	ドラッグアンドドロップ	75	情報検索
43	ブラウザ	76	情報統合
44	プラグイン	77	制約条件
45	フレームワーク	78	生成
46	プロパティ	79	定義する
47	プロパティの値	80	電子ドキュメント
48	マウス	81	統計分析
49	マッピング	82	特徴度
50	マークアップ	83	文字列
51	メタデータ	84	保守
52	リソース	85	<b>目的語</b>
53	リンク	86	曖昧性

## TEXTO 2

Términos en español			
1	algoritmo	17	<b>Mixturas de Gaussianas</b>
2	<b>andamiaje social</b>	18	<b>modelo</b>
3	<b>aprendizaje</b>	19	<b>Modelos de Mixtura de Gaussianas</b>
4	aprendizaje incremental	20	objeto virtual
5	autoexploración	21	pcincipio
6	autonomía	22	Regresión
7	categorizar	23	robot
8	coste computacional	24	robot humanoide
9	directriz	25	robot móvil
10	estado sensorial futuro	26	robótica
11	estructura cognitiva	27	robótica del desarrollo
12	Gaussiana	28	<b>scaffolding social</b>
13	herramienta cognitiva	29	sistema de aprendizaje
14	imitación	30	solución
15	ingeniería tradicional	31	substrato representacional
16	interacción	32	tarea

Términos en japonés			
1	MLDA	24	確率文法
2	mMLDA	25	学習
3	multilayered MLDA	26	関係性
4	Multimodal Latent Dirichlet Allocation	27	記号
5	インタラクション	28	記号接地問題
6	カテゴリ	29	教示発話
7	カテゴリ分類	30	計算モデル
8	シーン	31	自然言語処理
9	タスク	32	上位層
10	ヒューマノイド	33	人工知能
11	マルチモーダル情報	34	生起順
12	マルチモーダル潜在的ディレクトリ配分法	35	遷移確率
13	モデル	36	掃除タスク
14	ロボット	37	相互依存関係
15	ロボット技術	38	相互情報量
16	ロボティクス	39	多層マルチモーダル潜在的ディレクトリ配分法
17	音韻ラベル	40	入力パターン
18	下位概念	41	把持行動
19	下層	42	物体概念
20	階層的カテゴリ分類	43	物体認識
21	階層的カテゴリ分類	44	文脈
22	概念クラス	45	未観測情報
23	確率的知識表現		

### TEXTO 3

Términos en español			
1	abstracción	29	modelo operacional
2	adaptabilidad	30	multiagente
3	agente	31	parámetro
4	aplicación	32	predicción cualitativa
5	asistencia computacional	33	predicción cuantitativa
6	científico de computación	34	procedimiento
7	comportamiento	35	regla
8	condición emergente	36	representación
9	consistencia	37	semántica
10	coste	38	simulación
11	diseño	39	Simulación basada en Agentes
12	distribución	40	<b>Simulación basada en Sistemas Multiage</b>
13	E-commerce	41	Simulación de Agentes
14	elemento	42	simulación multiagente
15	exploración	43	simulación numérica
16	exploración manual	44	Simulación soportada por Agentes
17	fiabilidad incorporada	45	sistema
18	fórmula	46	sistema complejo
19	habilidad cognitiva	47	sistema de agente
20	heurística	48	Sistema de Simulación basado en Agentes
21	ingeniería eléctrica	49	Sistema Multiagente
22	inteligencia artificial	50	sistema social
23	interfaz	51	software
24	logística	52	solución abierta
25	<b>MASB</b>	53	solución adaptativa
26	mecatrónica	54	solución flexible
27	modelador	55	tipo de uso
28	modelo	56	variable

Términos en japonés			
1	CG	21	開発
2	POSデータ	22	基礎的検討
3	エージェント	23	機械学習
4	システム	24	距離値
5	シミュレーション	25	顧客
6	シミュレータ	26	自然言語処理
7	ステップ	27	自立性
8	セル	28	実用性
9	データマイニング	29	初期配置
10	パラメータ	30	将棋倒し
11	フィールド	31	人工社会
12	マルチエージェント	32	人工知能
13	マルチエージェントシミュレーション	33	人的被害
14	安全性	34	相互作用
15	移動方向	35	知的システム
16	遺伝アルゴリズム	36	繁盛時
17	音声認識	37	避難シミュレーション
18	可動障害物	38	避難者
19	画像認識	39	誘導灯
20	画像認識システム	40	漏斗状

#### TEXTO 4

Términos en español			
1	abstrucción	26	modelo matemático
2	algoritmo de aprendizaje	27	Modelo mediante SBRDs
3	área de aplicación	28	Red Neuronal
4	caja blanca	29	red semántica
5	caja gris	30	Redes de Base Radial
6	comportamiento no lineal	31	regla difusa
7	conjunto difuso	32	Reglas de Aprendizaje
8	dato de entrada	33	SBRD
9	dato de salida	34	SI-ENTONCES
10	estructura basada en reglas de los Sistemas Difusos	35	simulación numérica
11	etiquetas de formato	36	Sistema Basado en Reglas Difusas
12	funciones de pertenencia	37	sistema biológica
13	granularidad	38	sistema dinámico
14	interfaz	39	término difuso
15	lenguaje natural	40	término lingüístico
16	lógica difusa	41	tiempo real
17	metodología inteligente	42	vaguedad
18	modelado	43	variable
19	modelado de caja blanca	44	variable de entrada
20	modelado de caja gris	45	variable de salida
21	modelado de caja negra	46	variable difusa
22	modelado de datos	47	variable lingüística
23	modelado de sistema	48	variable numérica de entrada
24	modelo aproximado	49	variable numérica de salida
25	modelo cualitativo		

Términos en japonés			
1	<b>antecedent part</b>	5	<b>fuzzy rule</b>
2	<b>AならBである</b>	6	<b>Fuzzy Set</b>
3	<b>consequent part</b>	7	<b>If A then B</b>
4	<b>fuzzy inference</b>	8	<b>Self-Organizing Fuzzy Neural Network</b>

Términos en japonés			
9	<b>SOFNN</b>	31	広がり
10	<b>Takagi-Sugeno-Kang fuzzy model</b>	32	高木・菅野ファジィ推論モデル
11	<b>TSK fuzzy model</b>	33	次元
12	<b>ガウシアン型関数</b>	34	自己組織化ファジィニューラルネットワーク
13	クリスプ集合	35	実数パラメータ
14	システム固定	36	実数値
15	ニューラルネットワーク	37	実数値
16	バイナリ	38	重み
17	パラメータ	39	出力
18	ファジィニューラルネットワーク	40	人工神経回路網
19	<b>ファジィルール</b>	41	推論ルール
20	<b>ファジィ集合</b>	42	線形結合関数
21	<b>ファジィ推論</b>	43	前件部
22	ファジィ制御技術	44	全体集合
23	ファジィ理論	45	代数積
24	ファジィ制御	46	度合い
25	プロダクションシステム	47	特性関数
26	<b>ベル型関数</b>	48	部分集合
27	メンバーシップ関数	49	変数
28	基本概念	50	要素
29	教師学習機能	51	論理積
30	<b>後件部</b>		

## TEXTO 5

Términos en español			
1	agente externo	33	estímulos externos
2	<b>algoritmo de aprendizaje</b>	34	estructura multicapa
3	Aprendizaje "Adaptativo"	35	feedback
4	aprendizaje no supervisado	36	<b>feedforward</b>
5	aprendizaje supervisado	37	función arcotangente
6	Autoorganización	38	función de error
7	Backpropagation	39	<b>función Escalón</b>
8	capa	40	función identidad
9	capa de entrada	41	función sigmoideal
10	capa de salida	42	función umbral
11	capa oculta	43	gradiente
12	capacidad autoorganizativa	44	hacia delante
13	capacidad de aprendizaje	45	herencia genética
14	computador digital	46	hiperplano
15	computador neuronal	47	instante
16	<b>computador secuencial</b>	48	intensidad de las conexiones
17	<b>condición inicial</b>	49	intensidad de uniones
18	conexión sináptica	50	límite de saturación
19	conjunto borroso	51	lineal
20	constante del sistema	52	lógica borrosa
21	dato de entrada	53	magnitud
22	dato de salida	54	matriz de pesos
23	derivada positiva	55	Memoria Asociativa
24	destrucción	56	solución abierta
25	elemento de proceso	57	solución adaptativa
26	entrada	58	solución flexible
27	entrenamiento	59	tipo de uso
28	<b>Escalón</b>	60	variable
29	estado de activación	61	Mixta
30	estado de equilibrio	62	modelado
31	estado de excitación	63	modelo
32	estado de reposo	64	modelo analítico

65	modelo biológico	89	red
66	multicapa	90	red binaria
67	neurona	91	Red Neuronal
68	no lineal	92	Red neuronal artificial
69	número de capas	93	Red Neuronal Celular
70	Operación en tiempo real	94	Red Recurrente
71	operador	95	regla de decisión
72	operador no lineal	96	Reglas de Aprendizaje
73	<b>ordenador secuencial</b>	97	salida
74	parámetros internos	98	sigmoidal
75	patrón	99	silicio
76	patrón de entrada	100	sinapsis
77	patrón de salida	101	sistema de educaciones
78	perceptrón	102	sistema dinámico discreto
79	perceptrón multicapa	103	sistema neuronal
80	peso	104	sistema neuronal artificial
81	peso de conexión	105	sistema secuencial
82	peso de la red	106	suma ponderada
83	peso sináptico	107	Tolerancia a fallos
84	plano	108	transformación
85	Planta	109	umbral
86	Propagación hacia atrás	110	valor de activación
87	Propagación hacia atrás del error	111	vector de entrada
88	reconocimiento de patrones	112	vector de prueba

<b>Términos en japonés</b>			
1	<b>SOFNN</b>	30	<b>ニューロン</b>
2	<b>Back Propagation法</b>	31	<b>パターン</b>
3	<b>BP法</b>	32	パターン認識装置
4	<b>EX.OR</b>	33	パターン認識装置
5	<b>fully connected neural network model</b>	34	パターン分類
6	<b>Linear Transfer function</b>	35	バックプロパゲーション
7	<b>Log Sigmoid Transfer Function</b>	36	バックプロパゲーション訓練
8	<b>multi layered neural network model</b>	37	バックプロパゲーション法
9	<b>neuron</b>	38	パルス
10	<b>perceptron</b>	39	パーセプトロン
11	<b>Sigmoid関数</b>	40	フィードバック結合
12	<b>supervised learning</b>	41	フィードフォワード
13	synapse	42	フィードフォワードネットワーク
14	<b>S字型関数</b>	43	モデル
15	<b>unsupervised learning</b>	44	<b>モーメンタム</b>
16	アンダーフィッティング	45	ユニット
17	アーキテクチャ	46	ローカル
18	オーバーフィッティング	47	安定化定数
19	グローバル	48	因数
20	しきい値	49	隠れ層
21	シグモイド	50	可変
22	<b>シグモイド関数</b>	51	階層構造
23	シナプス	52	<b>階層構造ニューラルネットモデル</b>
24	シナプスウェイト	53	外部
25	シナプス結合	54	学習
26	シナプス結合強度	55	学習アルゴリズム
27	スカラー	56	学習形態
28	ニューラルネット	57	学習比
29	ニューラルネットワーク	58	学習法式

59	完全結合型	105	静止状態
60	<b>慣性項</b>	106	積
61	関数近似	107	線形
62	<b>逆誤差伝播則</b>	108	<b>線形伝達関数</b>
63	<b>教師あり学習</b>	109	線形分離可能の条件
64	<b>教師信号</b>	110	線形分離可能の条件
65	<b>教師無し学習</b>	111	選定
66	興奮状態	112	前抑制
67	興奮性	113	層
68	訓練	114	層数
69	結合の重み	115	相互結合
70	結合強度	116	<b>相互結合ニューラルネットモデル</b>
71	結合係数	117	総入力
72	結合路固有	118	多層
73	誤差	119	<b>対数シグモイド伝達関数</b>
74	誤差表面	120	単位ステップ関数
75	誤差評価関数	121	単純パーセプトロン
76	勾配ベクトル	122	単入力
77	再帰的	123	知能化
78	最急降下法	124	中間層
79	最適数	125	適応学習比
80	細胞体	126	伝達関数
81	自乗和	127	伝播
82	自由度	128	電位
83	軸索	129	電気パルス
84	実数パラメータ	130	動作原理
85	写像	131	内臓
86	樹状突起	132	内部状態
87	収束計算	133	入力
88	修正値	134	入力パターン
89	重み	135	入力ユニット
90	出力	136	入力信号
91	出力ユニット	137	入力層
92	出力層	138	入力端子
93	出力端子	139	脳神経系
94	出力値	140	<b>排他的論理和</b>
95	順方向	141	非線形
96	初期条件	142	微分
97	情報処理素子	143	評価基準
98	神経系	144	分類対象
99	<b>信号</b>	145	並列処理
100	信号伝達	146	変化
101	信号伝播	147	変数
102	振動	148	模擬
103	人工ニューロン	149	有限界
104	制約	150	抑制性

Anexo 4. Términos registrados en el glosario

	Término en japonés	Término en español	
1	エージェント	agente	1
2	外部	agente externo	2
3	学習アルゴリズム	algoritmo de aprendizaje	3
4	学習	aprendizaje	4
5	<b>unsupervised learning</b>	aprendizaje no supervisado	5
6	教師無し学習		
7	<b>supervised learning</b>	aprendizaje supervisado	6
8	教師あり学習		
9	<b>Back Propagation法</b>	Backpropagation Propagación hacia atrás del error	7
10	<b>BP法</b>		
11	<b>バックプロパゲーション法</b>		
12	<b>逆誤差伝播則</b>		8
13	層	capa	9
14	入力層	capa de entrada	10
15	出力層	capa de salida	11
16	隠れ層	capa oculta	12
17	中間層		
18	カテゴリ分類	categorizar	13
19	初期条件	condición inicial	14
20	シナプス結合	conexión sináptica	15
21	<b>Fuzzy Set</b>	conjunto difuso	16
22	<b>ファジィ集合</b>		
23	定義する	definir	17
24	記述する	describir	18
25	情報処理素子	elemento de proceso	19
26	<b>neuron</b>	neurona	20
27	<b>ニューロン</b>		21
28	入力	entrada	22
29	訓練	entrenamiento	23
30	スキーム	esquema	24
31	興奮状態	estado de excitación	25
32	静止状態	estado de reposo	26
33	階層構造	estructura multicapa	27
34	開発	exploración	28
35	フィードフォワード	feedforward	29
		hacia delante	30
36	誤差評価関数	función de error	31
37	<b>シグモイド関数</b>	función sigmoïdal	32
38	勾配ベクトル	gradiente	33
39	ツール	herramienta	34
40	HTML	HTML	35
41	識別子	identificador	36
42	人工知能	inteligencia artificial	37
43	結合強度	intensidad de uniones	38
		Intensidad de las conexiones	39
44	インタラクション	interacción	40
45	記述言語	lenguaje de descripción	41
46	XML構文	lenguaje XML	42
47	線形	lineal	43
48	ファジィ理論	lógica difusa	44
49	機械	máquina	45
50	フレームワーク	marco	46

51	メタデータ	metadato	47
52	モデル	modelo	48
53	マルチエージェント	multiagente	49
54	多層	multicapa	50
55	非線形	no lineal	51
56	層数	número de capas	52
57	オントロジ	ontología	53
58	<b>OWL</b>	OWL	54
59	<b>Web Ontology Language</b>		
60	パラメータ	parámetro	55
61	<b>信号</b>	patrón	56
62	<b>パターン</b>		
63	入力パターン	patrón de entrada	57
64	<b>perceptron</b>	perceptrón	58
65	<b>パーセプトロン</b>		
66	単純パーセプトロン		
67	重み	peso	59
68	結合の重み	peso de conexión	60
69	出力値	peso de la red	61
70	シナプスウェイト	peso sináptico	62
71	バックプロパゲーション	propagación hacia atrás	63
72	<b>RDF</b>	RDF	64
73	<b>Resource Description Framework</b>		
74	<b>RDF Scheme</b>	RDFS RDF Schema	65
75	<b>RDF Vocabulary Description Language 1.0</b>		
76	<b>RDFS</b>		
77	<b>RDFスキーマ</b>	66	
78	<b>ニューラルネットワーク</b>	red neuronal	67
79	<b>ニューラルネット</b>	sistema neuronal artificial	68
80	<b>人工神経回路網</b>	Red neuronal artificial	69
81	<b>fuzzy rule</b>	regla difusa	70
82	<b>ファジィルール</b>		
83	制約条件	restricción	71
84	ロボット	robot	72
85	ヒューマノイド	robot humanoide	73
86	ロボティクス	robótica	74
87	出力	salida	75
88	セマンティック	semántica	76
89	サーバ	servidor	77
90	<b>AならBである</b>	SI-ENTONCES	78
91	<b>If A then B</b>		
92	シグモイド	sigmoidal	79
93	シミュレーション	simulación	80
94	マルチエージェントシミュレーション	simulación multiagente	81
95	シナプス	sinapsis	82
96	システム	sistema	83
97	神経系	sistema neuronal	84
98	タスク	tarea	85
99	変化	transformación	86
100	変数	variable	87
101	入力データ	vector de entrada	88
102	セマンティックweb	web semántica	89