

LA ESTRUCTURA DISCURSIVA COMO CRITERIO DE EVALUACIÓN DE TRADUCCIONES AUTOMÁTICAS: UNA PRIMERA APROXIMACIÓN

MARINA FOMICHEVA¹

IRIA DA CUNHA FANEGO^{1,2,3}

GERARDO SIERRA MARTÍNEZ¹

¹Universidad Nacional Autónoma de México, México D.F.

²Universitat Pompeu Fabra, Barcelona

³Université d'Avignon et des Pays de Vaucluse, Avignon

RESUMEN

El objetivo del presente trabajo es evaluar dos sistemas de traducción automática mediante estrategias discursivas. Para ello, se compiló un corpus de textos especializados bilingües español-inglés. Los textos en español fueron traducidos al inglés por dos sistemas de traducción automática, uno con base lingüística y otro con base estadística. Se llevó a cabo la anotación discursiva de todos los textos del corpus. Finalmente, se compararon las estructuras discursivas de los textos originales en español con las respectivas estructuras de sus traducciones humanas y automáticas en inglés, para evaluar estas últimas. Los resultados de este estudio preliminar demuestran que la estructura discursiva principal del texto original debe mantenerse en el texto traducido y parecen sugerir que podría ser posible optimizar sistemas de traducción automática estadísticos mediante estrategias discursivas.

Palabras clave: traducción automática, evaluación, análisis del discurso, estructura discursiva

ABSTRACT

The present study aims to evaluate two systems of machine translation with discourse strategies. We made a corpus of bilingual specialized texts in Spanish and English. The texts were translated by two machine translation systems, one of them linguistic and the other one statistical. The rhetorical annotation of all the texts of the corpus was done. Finally, the rhetorical structures of the source texts were compared with the respective structures of human and machine translations, in order to evaluate the latter. The results of this preliminary study show that, basically, the rhetorical structure of the source text must be maintained in the target text and that it could be possible to optimize statistical machine translation systems by using discourse strategies.

Keywords: machine translation, evaluation, discourse analysis, rhetorical structure

1. INTRODUCCIÓN

La traducción automática (TA) es un área de estudio relativamente reciente en la que todavía existen muchos problemas sin resolver. Los sistemas de TA más utilizados en la actualidad son estadísticos o basados en reglas (Hutchins y Somers, 1992; Wilks, 2009), aunque ya ha habido intentos de construir sistemas híbridos que combinen estos dos métodos (Yu Chen y Eisele, 2010), ya que cada uno de ellos tiene ventajas y desventajas. Es necesario establecer criterios y métodos de evaluación que nos permitan detectar cuáles son sus carencias, a qué se deben, y qué es lo que podemos mejorar. La evaluación de TA puede ser manual y automática. Hutchins y Somers (1992) mencionan que para la evaluación manual se toman en cuenta tres criterios básicos: a) fidelidad y precisión, b) inteligibilidad y claridad, y c) estilo. Esta evaluación es cara y, en la mayoría de los casos, no reutilizable. La evaluación

automática tiene la ventaja de ser rápida, replicable y objetiva (Mayor et al., 2009). Sin embargo, el sistema más utilizado actualmente, BLEU (Papineni et al., 2002), basado en la comparación de coocurrencias de n-gramas entre una traducción automática y una o varias traducciones de referencia realizadas por humanos, ha recibido numerosas críticas. Mayor et al. (2009) afirman que no caracteriza de manera adecuada la calidad absoluta de la traducción ni la coherencia global. En cambio, en traductología sí existen trabajos sobre la calidad y la evaluación de la traducción en relación con el análisis del discurso (Larose, 1989). Nosotros consideramos que una de las perspectivas lingüísticas para acercarse al problema de la calidad en la TA podría ser el análisis discursivo, que permitiría evaluar la coherencia del texto. Precisamente una de las carencias de la mayoría de los sistemas de TA existentes es que traducen oración por oración, sin considerar el nivel discursivo, lo cual afecta a la calidad de la traducción (Wilks, 2009).

Los objetivos del presente trabajo son: a) realizar un estudio preliminar para averiguar de qué manera el hecho de conservar o modificar la estructura discursiva del texto original en el texto traducido afecta la calidad de la traducción y b) establecer algunos criterios para evaluar sistemas de TA a partir de la comparación de las estructuras discursivas del texto original (TO) y el texto traducido (TT).

2. MARCO TEÓRICO

La *Rhetorical Structure Theory* (RST) de Mann y Thompson (1988) es una teoría que permite describir la estructura del texto en términos funcionales, con base en las relaciones discursivas que se establecen entre las unidades del mismo. El texto posee una estructura jerárquica, en la cual cada unidad tiene una función determinada. De esta manera se pone de relieve el papel comunicativo de la estructura discursiva del texto. Según la RST, un texto puede segmentarse en unidades discursivas elementales o mínimas (EDUs), que pueden ser de dos tipos: núcleos (elementos gobernantes) y satélites (elementos que dependen de los núcleos y aportan cierta información sobre ellos). Entre estas unidades pueden existir relaciones multinucleares (ej. Lista, Secuencia, Contraste, etc.) y núcleo-satélite (ej. Condición, Resultado, Causa, etc.).

La RST tiene ciertas ventajas para los propósitos de esta investigación: ya se ha aplicado con éxito en la lingüística computacional, y, en particular, en la generación automática de textos, la cual constituye una de las etapas de la TA; se ha utilizado para realizar estudios comparativos de diversas lenguas (Taboada y Mann, 2006), y permite dar cuenta de características del texto relevantes para la evaluación de la traducción: la coherencia y el papel comunicativo de la estructura discursiva. En cuanto a la aplicación de la RST en TA, se han realizado pocos estudios (Ghorbel et al., 2001; Rösner, 1993; Marcu et al., 2000).

3. METODOLOGÍA

3.1 Herramientas de TA

En este trabajo, por medio del análisis de la estructura discursiva, comparamos las traducciones producidas por dos sistemas de traducción automática: Traductor de Google¹ y Lucy LT KWIK Translator². El primero es un sistema de TA basado en técnicas estadísticas que genera traducciones a partir de patrones hallados en un gran número de textos bilingües paralelos. El segundo es un sistema de TA basado en reglas lingüísticas que posee una estructura modular de gramáticas computacionales y lexicones.

3.2 Compilación del corpus

Para realizar esta investigación compilamos un corpus bilingüe paralelo de 15 textos especializados de cinco ámbitos: psicología, procesamiento del lenguaje natural, ingeniería sísmica, economía y derecho. El corpus contiene resúmenes de artículos de investigación en español y en inglés, y las traducciones automáticas al inglés de los resúmenes originales³. Para la compilación del corpus utilizamos artículos publicados entre 2006 y 2010 en revistas científicas. El corpus contiene en total 9783 palabras y 627 EDUs.

3.3 Anotación discursiva del corpus

Para analizar la estructura discursiva de los textos realizamos la anotación del corpus con la metodología de la RST en tres etapas: segmentación, detección de relaciones y construcción de árboles discursivos. En la línea de los trabajos de Tofilosky et al. (2009) y da Cunha e Iruksieta (2010), consideramos como EDUs los elementos del texto que: a) presenten claramente una relación discursiva y b) contengan un verbo conjugado, un gerundio o un infinitivo. No consideramos como EDUs las cláusulas subordinadas de relativo y de complemento. Utilizamos la lista extendida de relaciones de Mann (2005). Para construir los árboles discursivos utilizamos la RSTTool⁴ (O'Donnell, 2000). La Figura 1 muestra un ejemplo de árbol de un fragmento del corpus.⁵

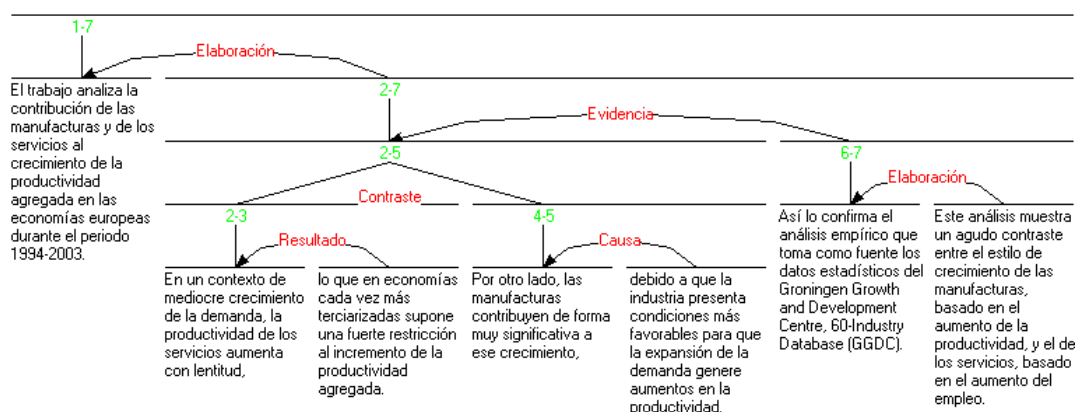


Figura 1. Árbol discursivo

3.4 Comparación de la estructura discursiva

Para averiguar de qué manera el hecho de conservar o modificar la estructura discursiva afecta la calidad de la traducción, comparamos la estructura de los textos originales con las traducciones humanas, evaluando las modificaciones realizadas desde el punto de vista traductológico. Después comparamos la estructura discursiva de los textos originales con la de las traducciones automáticas realizadas por los dos sistemas de TA. Nos fijamos solo en el aspecto discursivo de la traducción; las diferencias a nivel sintáctico o léxico se analizan solo si afectan la estructura discursiva.

4 RESULTADOS

4.1 Comparación de la estructura discursiva de TO y los TT producidos por humanos

La segmentación discursiva en los TT cambia un 9.4% en relación con la segmentación de los TO. Las relaciones discursivas varían en un 5.7% y la nuclearidad en un 3.5%. La Tabla 1 muestra los resultados.

Texto	Segmentación		Relaciones	Nuclearidad
	Añadida	Eliminada		
Psicología: Texto 1	0	0	0	0
Psicología: Texto 2	1	1	1	1
Psicología: Texto 3	3	2	0	0
PLN: Texto 1	0	0	0	0
PLN: Texto 2	0	0	1	1
PLN: Texto 3	0	0	0	0
Ingeniería: Texto 1	0	1	1	1
Ingeniería: Texto 2	1	1	0	0
Ingeniería: Texto 3	1	1	0	0
Economía: Texto 1	0	0	0	0
Economía: Texto 2	0	1	1	1
Economía: Texto 3	0	0	1	1
Derecho: Texto 1	0	0	1	0
Derecho: Texto 2	0	1	2	0
Derecho: Texto 3	0	1	0	0
Total	6	9	8	5
Total absoluto	15		8	5

Tabla 1. Análisis cuantitativo TO-TT producidos por humanos

Las modificaciones que realiza el traductor no siempre están justificadas, porque la traducción no se hizo por un traductor profesional, sino por el autor del artículo. Por este motivo, hemos cuantificado las modificaciones del TT que, a nuestro juicio, no tendrían por qué haberse realizado. Así, obtendríamos que la segmentación cambia en un 7.5% de los casos, las relaciones en un 1.4% y la nuclearidad en un 0.7%. La Tabla 2 resume estos resultados.

	Segmentación	Reaciones	Nuclearidad
Modificaciones justificadas	12	2	1
Modificaciones injustificadas	3	6	4
Total	15	8	5

Tabla 2. Modificaciones justificadas en injustificadas TO-TT

Los cambios en la segmentación en la traducción humana se deben a las modificaciones de la estructura sintáctica. En inglés las oraciones suelen ser más breves y los mecanismos de cohesión son diferentes. En estos casos los cambios de segmentación son justificados, ya que no afectan de manera significativa la estructura discursiva y se deben a diferencias lingüísticas entre la lengua de partida y la de llegada. La modificación de la segmentación en las traducciones humanas en ocasiones conlleva una alteración de la relación. En estos casos los cambios en la segmentación casi siempre son injustificados.

Los cambios en las relaciones discursivas son injustificados prácticamente en todos los casos. Se deben a omisiones o sustituciones de marcadores discursivos. Esto refleja cierta arbitrariedad de la traducción humana, que puede afectar la calidad del TT a nivel discursivo.

El ejemplo 1 muestra un cambio de la relación y también un cambio en la nuclearidad:

1. [Mi posición es favorable a la adhesión turca,] [pero pienso que deben valorarse previamente las implicaciones y la forma de afrontar esa ampliación.]

[While I support Turkish adhesion,] [I also suggest prior evaluation of the implications, as well as of the actions necessary to prepare Turkey's EU membership.]

Al sustituir el marcador discursivo "pero" (que indica Antítesis) por el marcador "while" (que marca Concesión), se invierte la estructura del fragmento y se produce un cambio de la nuclearidad.

En general, los cambios de nuclearidad se producen porque las relaciones núcleo-satélite se transforman en relaciones multinucleares, o a la inversa. Estos cambios casi siempre evidencian un error en la traducción, ya que la estructura jerárquica del TO refleja el grado de importancia que le da el autor a cada fragmento.

4.2 Comparación de la estructura discursiva de los TO y los TT producidos por sistemas automáticos

4.2.1 Traductor de Google

La segmentación discursiva del TT automático cambia un 5% en relación con la segmentación del TO. Las relaciones del TT varían en un 7.8% y la nuclearidad en un 4.2%. La Tabla 3 muestra los resultados.

Texto	Segmentación		Relaciones	Nuclearidad
	Añadida	Eliminada		
Psicología: Texto 1	0	0	0	0
Psicología: Texto 2	0	0	1	1
Psicología: Texto 3	0	2	0	0
PLN: Texto 1	0	0	0	0
PLN: Texto 2	0	1	1	0
PLN: Texto 3	0	0	1	0
Ingeniería: Texto 1	3	0	0	0
Ingeniería: Texto 2	0	0	4	3
Ingeniería: Texto 3	0	0	1	1
Economía: Texto 1	0	0	0	0
Economía: Texto 2	0	1	0	0
Economía: Texto 3	0	0	0	0
Derecho: Texto 1	0	0	1	0
Derecho: Texto 2	0	1	2	1
Derecho: Texto 3	0	0	0	0
Total	3	5	11	6
Total absoluto	8		11	6

Tabla 3. Análisis cuantitativo TO-TA del Traductor de Google

Los cambios en la segmentación se deben a omisiones o sustituciones de unidades léxicas en el TT, que normalmente evidencian errores de traducción. Sin embargo, en ocasiones, en la TA estadística se producen cambios en la segmentación similares a los de la traducción humana. Creemos que estos cambios se deben a que el sistema modifica la estructura sintáctica a partir de patrones extraídos estadísticamente de la Web. Por ejemplo:

2. [En relación a la Potencia, ésta fue alta para todas las variables estudiadas,] [presentando un valor de 1] [al finalizar la investigación.]

a. [In relation to Power, it was high for all variables studied, with a value of 1 at the end of the investigation].

b. [In relation to Power, it was high for all variables studied, with a value of 1 at the end of the research.]

El fragmento 2a es la traducción del Traductor de Google y el 2b es la traducción del autor. En ambos casos se sustituye el gerundio "presentando" por la conjunción "with", y la estructura "al + infinitivo" por la frase preposicional "at the end [...]", de la misma manera que ocurre en la traducción humana.

Como en la traducción humana, los *cambios de relación discursiva* en las TA de este sistema son siempre injustificados y se deben a omisiones y sustituciones de unidades léxicas; los *cambios de nuclearidad* siempre se deben a un cambio de relación.

4.2.2 Lucy LT KWIK Translator

La segmentación discursiva del TT automático cambia un 3.1% en relación con la segmentación del TO. Las relaciones del TT varían en un 2.1% y la nuclearidad en un 1.4%. La Tabla 4 muestra los resultados.

Texto	Segmentación		Relaciones	Nuclearidad
	Añadida	Eliminada		
Psicología: Texto 1	0	0	0	0
Psicología: Texto 2	0	1	0	0
Psicología: Texto 3	0	0	0	0
PLN: Texto 1	0	0	0	0
PLN: Texto 2	0	0	0	0
PLN: Texto 3	0	0	1	0
Ingeniería: Texto 1	1	0	1	1
Ingeniería: Texto 2	0	0	1	1
Ingeniería: Texto 3	0	0	0	0
Economía: Texto 1	1	0	0	0
Economía: Texto 2	0	0	0	0
Economía: Texto 3	0	0	0	0
Derecho: Texto 1	0	0	0	0
Derecho: Texto 2	0	1	0	0
Derecho: Texto 3	0	1	0	0
Total	2	3	3	2
Total absoluto	5		3	2

Tabla 4. Análisis cuantitativo TO-TA de Lucy LT KWIK Translator

Tanto los *cambios en la segmentación*, como los *cambios en las relaciones y nuclearidad*, que provocan una traducción inadecuada, se deben a errores en el análisis morfosintáctico del TO. Por ejemplo:

3. [Para ello, se estiman las ductilidades máximas [...],] [y se determinan factores de reducción de resistencia por ductilidad] [...].

[For that, the maximum ductility [...] is estimated], [and it is determined by factors of reduction of endurance/resistance for ductility] [...].

Al alterarse la estructura sintáctica de la oración (cambio de sujeto en la segunda cláusula), la relación multinuclear de Secuencia del TO se convierte en Elaboración (núcleo-satélite) en el TT.

4.2.3 Diferencias entre el Traductor de Google y Lucy LT KWIK Translator

La comparación de los árboles discursivos de las traducciones de estos dos sistemas permite dar cuenta de las diferencias entre métodos lingüísticos y estadísticos. La variación en la estructura discursiva en general es menor en las traducciones de Lucy LT KWIK Translator, ya que las traducciones producidas son en muchos casos literales. Precisamente, supera las traducciones del Traductor de Google por no realizar transformaciones injustificadas que este produce con base en patrones estadísticos.

El problema de la TA basada en reglas es su falta de robustez. Si hay un mínimo error en el TO, el sistema no puede traducirlo correctamente. En cambio, la TA estadística es capaz de resolver este problema pero no otros derivados de la carencia de una base lingüística.

En la Tabla 5 se observan las cinco causas de modificaciones de la estructura discursiva detectadas en nuestro corpus, con sus respectivas frecuencias de aparición.

Tipo de modificación	Traductor de Google (%)	Lucy LT KWIK Translator (%)
Cambio estructura sintáctica	43.7	75
Cambio de marcador discursivo	12.5	25
Omisión de marcador discursivo	18.7	0
Cambio de verbo	6.25	0
Omisión de verbo	18.7	0

Tabla 5. Causas de modificación de la estructura discursiva TO-TT

5. CONCLUSIONES

En este trabajo presentamos una comparación de las estructuras discursivas de resúmenes de artículos científicos en español con traducciones al inglés realizadas por los autores de los artículos y traducciones automáticas realizadas por un sistema estadístico (Traductor de Google) y un sistema basado en reglas (Lucy LT KWIK Translator).

Los resultados de la comparación indican que la variación de la estructura discursiva del TO en las traducciones humanas no es significativa. Esta se debe a diferencias lingüísticas entre la lengua fuente y la lengua meta, se refleja en la segmentación y prácticamente no afecta las relaciones discursivas ni la nuclearidad.

La comparación en cuanto a las estructuras discursivas de las traducciones automáticas permite evaluar los sistemas de TA desde el punto de vista del análisis discursivo. Lucy LT KWIK Translator, al ofrecer una traducción prácticamente literal del TO, en general, mantiene la estructura discursiva, pero no realiza modificaciones necesarias para reflejar las características propias de la lengua meta. En cambio, en las traducciones del Traductor de Google la estructura discursiva del TO se modifica de manera significativa. En algunos casos esta modificación corresponde a los cambios realizados en la traducción humana, lo cual ilustra las ventajas de los métodos estadísticos frente a los lingüísticos. Sin embargo, en muchas ocasiones el cambio en la estructura discursiva del TO en las traducciones automáticas se debe a omisiones y sustituciones injustificadas de marcadores discursivos. Así, el TT pierde coherencia. Creemos que sería posible optimizar sistemas de TA estadísticos por medio de estrategias del discurso.

En el futuro pensamos realizar experimentos con un corpus más amplio, analizar traducciones humanas y automáticas del inglés al español, analizar otro par de lenguas y optimizar un sistema de TA estadístico mediante estrategias discursivas.

NOTAS

1. <http://translate.google.com.mx/?hl=es&tab=wT#>
2. http://www.lucysoftware.de/cms/front_content.php?idcat=45&changelang=3
3. Se comprobó que los autores de los artículos eran hispanohablantes, lo cual implica que los resúmenes se redactaron en español y se tradujeron al inglés.
4. <http://www.wagsoft.com/RSTTool/>
5. Todos los ejemplos se extrajeron del corpus.

REFERENCIAS BIBLIOGRÁFICAS

- da Cunha, I., Iruskieta, M. 2010. "Comparing rhetorical structures of different languages: The influence of translation strategies". *Discourse Studies*, 12(5): 563-598.
- Ghorbel H., Ballim A., Coray G. 2001. "ROSETTA: Rhetorical and Semantic Environment for Text Alignment", en Rason, P., et al. (eds.) *Proceedings of Corpus Linguistics 2001*, 224-233.
- Hutchins W. J., Somers H. L. 1992. *An Introduction to Machine Translation*. London: Academic Press.
- Larose R. 1989. *Théories contemporaines de la traduction*. Québec: Presses de l'Université du Québec.
- Mayor A., Alegria I., Díaz de Ilarraza A. Labaka G., Lersundi M., Sarasola K. 2009. "Evaluación de un sistema de traducción automática basado en reglas o por qué BLEU sólo sirve para lo que sirve", *Procesamiento del Lenguaje Natural*, 43: 197-205.
- Mann W. C., Thompson S. A. 1988. "Rhetorical Structure Theory: Towards a functional theory of text organization", *Text*, 8(3): 243-281.
- Mann W. C. 2005. RST Web Site. www.sfu.ca/rst
- Marcu D., Carlson L., Watanabe M. 2000. "The automatic translation of discourse structures", en *Proceedings of the NACACL*: 9-17.
- O'Donnell M. 2000. "RSTTOOL 2.4 - A markup tool for rhetorical structure theory", en *Proceedings of the Natural Language Generation Conference*: 253-256.
- Papineni K., Roukos S., Ward T., Wei-Jing Zhu. 2002. "BLEU: a Method for Automatic Evaluation of Machine Translation", en *Proceedings of the 40th ACL*: 311-318.
- Rösner D. 1993. "Intentions, rhetoric, or discourse relations? A case from multilingual document generation", en *Proceedings of Workshop on Intentionality and Structure in Discourse Relations, ACL*, 106-109.
- Taboada M., Mann W. C. 2006. "Applications of Rhetorical Structure Theory", *Discourse Studies*, 8(4): 567-588.
- Tofilofsky M., Brooke J., Taboada M. 2009. "A syntactic and lexical-based discourse segmenter", en *Proceedings of the 47th ACL*.
- Wilks Y. 2009. *Machine translation: Its scope and limits*. Nueva York: Springer.
- Yu Chen, Eisele A. 2010. "Integrating a rule-based with a hierarchical translation system", en *Proceedings of LREC 2010*.