

Comunicación del conocimiento propio y relaciones discursivas en el género tesis

Paulina Meza - Universidad de La Serena
pmeza@userena.cl

Iria da Cunha - Universidad Nacional de Educación a Distancia
iriad@flog.uned.es

Rebut / Received: 20-11-17

Acceptat / Accepted: 23-5-18

Resum. Comunicació del coneixement propi i relacions discursives en el gènere tesi. Una de les principals dificultats que té un estudiant a l'escriure la seva tesi és expressar el coneixement propi. La nostra hipòtesi és que hi ha algunes funcions de la comunicació del coneixement propi (FCCP) en el gènere tesi que tenen una correspondència substancial amb les relacions discursives (RD) de la Teoria de l'Estructura Retòrica, més coneguda com Rhetorical Structure Theory (RST). L'establiment d'una correlació permetrà oferir als estudiants les marques discursives emprades amb freqüència per expressar les FCCP en aquest gènere i, així, ajudar-los a redactar adequadament les seves tesis. Plantegem com a objectiu determinar entre quines FCCPs i RD hi ha una correspondència substancial en el gènere tesi. Els resultats indiquen que hi ha correspondències rellevants entre algunes FCCP i RD, mentre que en altres casos no es donen aquestes correspondències o no són rellevants. A partir de les FCCP i RD correlacionades proposem dues estratègies complementàries per ajudar els estudiants a expressar el coneixement propi en les seves tesis.

Paraules clau: tesi, funcions de la comunicació del coneixement propi, relacions discursives, Rhetorical structure theory (RST).

Abstract. Communication of the author's own knowledge and discourse relations in the thesis genre. One of the greatest challenges for a college student, from any academic degree, is writing a thesis. In doing so, candidates must deal with the difficulties of expressing their own knowledge, for example, elaborating a research justification. In this work, our hypothesis is that there exist some functions in communicating one's knowledge (FCCPs) in the thesis genre that have a significant correspondence with the discourse relations (DRs) of the rhetorical structure theory (RST). This correlation would provide students with the discourse marks frequently used in the expression

of FCCPs in such genre and, thus, help them to write their thesis properly. In this context, the purpose of this study is to determine the point of significant correspondence between FCCPs and DRs in the thesis genre. Based on the correlation between FCCPs and DRs, we have proposed two complementary strategies to help students express their own knowledge in their thesis.

Keywords: thesis, functions of communicating one's knowledge, discourse relations, Rhetorical Structure Theory (RST).

1. Introducción

La tesis es un género que permite al estudiante progresar en su formación académica y, en el nivel más avanzado, dar pasos sólidos para ingresar en la comunidad científica. Sin embargo, concluir con éxito una tesis es muy complejo, tanto que un gran número de estudiantes necesitan muchos años para concluirla y otros no son capaces de hacerlo (da Cunha, 2016; Hidalgo y Passarella, 2009).

Un aspecto definitorio de la tesis es que constituye un medio a través del cual el estudiante no solo debe acreditar conocimiento, sino que también debe saber cómo dar cuenta de lo que ha investigado y de lo que conoce. Por tanto, para ingresar en una comunidad, debe ser capaz de comunicar el conocimiento propio y ha de hacerlo de acuerdo con las exigencias de su comunidad, demostrando, así, que domina la práctica discursiva académica (Meza, 2015).

Una capacidad indispensable a la hora de redactar un texto de estas características es hilar las diferentes ideas que se quieren expresar de una manera coherente y cohesionada. Además, dichas ideas deben expresar los distintos contenidos que son necesarios en el género en cuestión. Meza (2013) concluye que, en el género tesis, hay 40 tipos de contenidos que expresan aportaciones del autor del trabajo, esto es, las funciones de la comunicación del conocimiento propio (FCCP). Por ejemplo, detectar vacíos en el conocimiento, indicar objetivos, etc. Estos contenidos pueden expresarse de diversas maneras formalmente. Sin embargo, nuestra hipótesis es que hay ciertos contenidos de las FCCP cuya formulación es recurrente. Por ejemplo, el modo de definir conceptos seguirá ciertos patrones discursivos, que implican la selección de determinadas estructuras sintácticas y unidades léxicas. En muchas ocasiones, los estudiantes que se encuentran frente al reto de expresar este tipo de funciones en su tesis no disponen de recursos que les ayuden en esta tarea. Apenas existe investigación sobre este tema en la actualidad y consideramos que es un camino que debe ser explorado.

Para verificar nuestra hipótesis, hemos tomado como marco teórico la Teoría de la Estructura Retórica, más conocida como Rhetorical Structure Theory (RST) de Mann y Thompson (1988), una teoría de organización textual que permite segmentar un texto en unidades discursivas mínimas o segmentos discursivos, entre los cuales se establecen relaciones retóricas o discursivas (RD) de diferentes tipos, como serían 'causa', 'reformulación', 'resumen', etc. Así, consideramos que es posible establecer una

correspondencia sustancial entre algunas de las FCCP (Meza, 2013) que aparecen en el género tesis y algunas de las RD de la RST. Para detectar esta correspondencia, se observarán marcas discursivas superficiales que evidencien las RD, como por ejemplo marcadores o conectores discursivos, determinadas estructuras sintácticas, ciertas unidades léxicas, etc., en la línea de trabajos previos en este ámbito (da Cunha, San-Juan, Torres-Moreno, Cabré y Sierra, 2012; Iruskietia, Díaz de Ilarraza y Lersundi, 2015; Pardo y Nunes, 2008; Tofiloski, Brooke y Taboada, 2009; Taboada, 2016; entre otros).

Saber formular una relación discursiva es relativamente sencillo dado que, por nuestro conocimiento del mundo y por el dominio del lenguaje que tenemos, lo hacemos de manera más o menos intuitiva. Sabemos qué marcadores emplear, qué estructuras lingüísticas utilizar, cómo resaltar la información que nos interesa, etc. En cambio, expresar las distintas FCCP en una tesis no resulta tan evidente para un estudiante. Así, al establecer una correlación entre determinadas FCCP y RD en el género tesis, puede ofrecerse a los estudiantes algunas de las marcas discursivas y estructuras empleadas con frecuencia para expresar las FCCP en este género. De este modo, dispondrán de un recurso, inexistente hasta la fecha, que les ayude a redactar adecuadamente los diferentes contenidos que deben incorporar en sus tesis.

El objetivo principal de este trabajo es determinar entre qué FCCP y RD existe una correspondencia sustancial en el género tesis. El segundo objetivo es proponer estrategias discursivas para la redacción de las FCCP en dicho género.

En cuanto a la organización de este artículo, en el segundo apartado presentamos el estado de la cuestión, donde mencionamos algunos de los principales trabajos relacionados con el género tesis, con las funciones con las que un sujeto escritor se inscribe en su discurso y con la redacción asistida por ordenador. En el tercero, explicamos el marco teórico empleado, dividido en dos subapartados: las FCCP y la RST. En el cuarto, detallamos la metodología diseñada para realizar nuestra investigación. En el quinto, exponemos el análisis realizado y los principales resultados obtenidos. A partir de estos resultados, en el sexto apartado, planteamos algunas estrategias que pueden ayudar a los estudiantes a redactar las FCCP de sus tesis. Finalmente, en el séptimo apartado aportamos las conclusiones del estudio y mencionamos algunas líneas de trabajo futuro.

2. Estado de la cuestión

En este apartado mostramos algunos trabajos relacionados con el género tesis, con las funciones con las que un escritor se inscribe en su discurso y con la redacción asistida por ordenador.

2.1. El género tesis

La tesis ha sido estudiada bajo distintos enfoques y modelos. Sin embargo, a pesar de su valor fundamental y de la dificultad que reviste su realización, existen algunas

dimensiones asociadas a este género que no han sido abordadas. Por un lado, son pocos los trabajos que han considerado la forma en que el autor comunica el conocimiento, aspecto que constituye la esencia de un trabajo de este tipo (Meza, 2013, 2015; Meza y Sabaj, 2016). Por otro, no se ha analizado la estructura y las relaciones discursivas de la RST que existen en dicho género.

En los últimos años ha habido, en general, un creciente interés por la escritura académica y, en particular, son muchos los esfuerzos dirigidos a ayudar a los estudiantes en la producción de sus trabajos de graduación. Sin embargo, estos esfuerzos se han centrado, principalmente, en la estructura del trabajo, en los métodos de investigación y en los aspectos sociales influyentes en el proceso de escritura, pero no en cómo dar cuenta del conocimiento que se pretende aportar mediante la tesis.

Además, la mayoría de las propuestas existentes no se fundamentan en investigaciones empíricas (Botta, 2002; Eco, 2006; Sabino, 2003; entre otros). Los trabajos que sí se sustentan empíricamente se han centrado en el análisis de algunas secciones y no en tesis completas (Maulid, 2017; Samraj, 2008; Tapia y Burdiles, 2012; Wuttisrisiriporn, 2017), se han enfocado en el estudio de tesis en lengua inglesa (Bui, 2013; Eckel, 2011; Paltridge y Starfield, 2007; Samraj, 2008), consideran solo un grado académico (Bui, 2013; Eckel, 2011; Martínez 2012, 2015; Samraj, 2008) y, además, ninguno de ellos ofrece estrategias concretas para ayudar a los estudiantes en la escritura de sus trabajos.

2.2. Funciones con las que un sujeto escritor se inscribe en su discurso

La inscripción de la persona en el texto es una de las características principales del discurso académico, pues cada escritor debe, necesariamente, optar por alguna forma para presentarse en su discurso. A pesar de ello, la investigación sobre este tema es escasa y se ha desarrollado, principalmente, sobre la base del estudio del Artículo de Investigación Científica (AIC). Kuo (1999), por ejemplo, señala doce funciones con las que alguien se inscribe en un AIC; Harwood (2005) reconoce siete funciones de la inscripción del yo en textos académicos (Reportes de proyectos y AIC); Tang y John (1999) analizan ensayos de estudiantes y revelan seis roles del uso de la primera persona (*genre roles*).

Para el género tesis, las investigaciones son aún menos frecuentes. Un caso es el trabajo de Savio (2010), quien presenta distintos roles de los estudiantes: escritor, investigador y psicoanalista. Asimismo, destacamos la investigación de Meza (2013), en la cual nos basamos en este trabajo y que explicamos detalladamente más adelante.

2.3. Redacción asistida por ordenador

Hoy en día existen múltiples enfoques y trabajos relacionados con la enseñanza de la escritura. Para el español contamos con diversos manuales de referencia para la escritura especializada como, por ejemplo, los de Cassany (2007) y Montolío (2014). También se

encuentran trabajos recientes que se centran en la redacción del género tesis (da Cunha 2016; Venegas, Núñez, Zamora y Santana, 2015; entre otros).

Asimismo, en los últimos años se han desarrollado sistemas computacionales que ayudan a la redacción de textos. En el caso de la lengua española, un ejemplo es arText, un asistente en línea para la redacción de distintos géneros textuales de ámbitos especializados, dirigido a diferentes tipos de usuarios (especialistas, estudiantes o legos), que incorpora información textual, léxica y discursiva. Uno de los géneros que incluye es, precisamente, el trabajo de fin de grado, concretamente en el ámbito médico. También existen otros sistemas que detectan en textos generales escritos en español errores gramaticales, ortográficos e incluso algunos de estilo, como son los sistemas comerciales Stilus y el sistema Language Tool, el cual es multilingüe y de código abierto. Para la lengua inglesa existen asimismo sistemas para escribir textos académicos, como, por ejemplo, SWAN, CALeSE y eWritingPal. También se han desarrollado diversos sistemas que ayudan a corregir errores gramaticales y ortográficos en esta lengua, como los sistemas comerciales Acrolinx, Grammarly y Grammar Checker, entre otros.

3. Marco teórico

3.1. Funciones en la comunicación del conocimiento propio

Entendemos la comunicación del conocimiento como el diálogo que se produce en la práctica discursiva científica para hacer circular la información. Este diálogo se instancia en los textos mediante diversos recursos lingüísticos, que están restringidos por las particularidades del discurso académico. El intercambio comunicativo se realiza en las tesis mediante los recursos que el autor utiliza, por una parte, para inscribirse en su discurso y, por otra, para incluir en él a otros participantes, mediante mecanismos de citación (Meza, 2013, 2015). En esta investigación, nos centramos en las funciones con las que el escritor de tesis se inscribe en su tesis (las FCCP) y que, según los trabajos basados en corpus de Meza (2013), son las incluidas en la Tabla 1.

TABLA 1. FCCP INCLUIDAS EN MEZA (2013)

1. Presentar la investigación o aspectos de ella	22. Emitir juicios sobre un concepto o teoría
2. Indicar vacíos/necesidades en el conocimiento	23. Indicar acuerdo con un autor y/o su planteamiento
3. Problematicar	24. Sintetizar un planteamiento teórico.
4. Justificar la investigación	25. Complementar un planteamiento teórico
5. Indicar organización de la investigación	26. Comparar planteamientos teóricos
6. Aludir a la tesis misma anafóricamente	27. Aplicar teorías o investigaciones a otros contextos
7. Comparar lo propio con lo de otros	28. Proponer aspectos teóricos
8. Explicar un gráfico, cuadro o figura de otro	29. Destacar la importancia de un concepto o teoría
9. Explicar un gráfico, cuadro o figura propio	
10. Indicar dificultades	
11. Indicar proyecciones	

12. Indicar limitación	30. Indicar opción teórica
13. Superar limitación	31. Introducir conceptos
14. Indicar hipótesis	32. Exponer hallazgos de la investigación
15. Indicar objetivos	33. Ejemplificar los resultados y hallazgos de la investigación
16. Describir una metodología o instrumento propio	34. Describir material didáctico
17. Explicar procedimientos	35. Realizar interpretaciones o deducciones
18. Definir un concepto	36. Contradecir investigaciones previas
19. Ejemplificar un concepto o teoría	37. Corroborar investigaciones previas
20. Enmarcar la investigación o un concepto relevante para ella	38. Corroborar hipótesis de investigación
21. Explicar un concepto o teoría	39. Refutar hipótesis de investigación
	40. Indicar aportes

3.2. Rhetorical Structure Theory (RST)

La RST es una teoría descriptiva de organización del texto muy útil para describirlo caracterizando su estructura a partir de las relaciones que mantienen entre sí los segmentos discursivos. Estas relaciones pueden ser núcleo-satélite (un elemento aporta una información adicional al otro) y multinucleares (ambos elementos tienen la misma relevancia para los propósitos del autor). Ejemplos de relaciones núcleo-satélite son 'condición', 'causa', 'medio', 'propósito' y 'elaboración', mientras que ejemplos de relaciones multinucleares son 'lista', 'secuencia' y 'contraste'.

TABLA 2. RELACIONES DISCURSIVAS DE LA RST

Relaciones N-N	Relaciones N-S	
Conjunción	Antítesis	Fondo
Contraste	Alternativa	Interpretación
Disyunción	Capacitación	Justificación
Lista	Causa	Medio
Secuencia	Concesión	Motivación
Unión	Condición	No condicional
	Condición inversa	Preparación
	Comparación	Propósito
	Circunstancia	Reformulación
	Elaboración	Resultado
	Evaluación	Resumen
	Evidencia	Solución

El análisis con la RST consta de tres etapas: segmentación discursiva, detección de RD y construcción de árboles retóricos. En este trabajo, los criterios de segmentación discursiva son similares a los empleados en la construcción del *RST Spanish Treebank* (da Cunha, Torres-Moreno y Sierra, 2011). En cuanto a las RD, existen diversas clasificaciones: la clásica de Mann y Thompson de 24 RD (Mann y Thompson, 1988), la extendida de Mann y Thompson de 30 RD (Mann, 2005) y la clasificación de Marcu de 136 RD (Carlson, Marcu y Okurowski, 2001). En este trabajo hemos seleccionado la clasificación extendida (Mann, 2005) para la anotación de las RD de nuestro corpus. En la Tabla 2 se incluyen dichas RD. En la primera columna se muestran las RD multinucleares (N-N), mientras que en las dos siguientes columnas se recogen las RD núcleo-satélite (N-S).

En la Figura 1 se incluye un ejemplo de nuestro corpus analizado con la RST, mediante la interfaz de anotación RSTTool (O'Donnell, 2000).

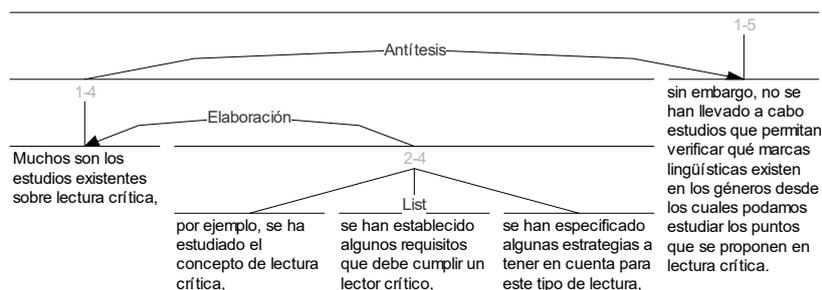


FIGURA 1. EJEMPLO DE ÁRBOL DISCURSIVO DE LA RST.

4. Metodología utilizada en el estudio

La metodología de esta investigación incluye diversas fases. En primer lugar, se conformó el corpus de trabajo. Partimos de un corpus inicial recopilado en Meza (2013), denominado corpus TeLing3N, que incluye 36 tesis de lingüística (12 de licenciatura, 12 de máster y 12 de doctorado) de universidades chilenas pertenecientes al Consejo de Rectores de Universidades Chilenas. El número total de palabras del corpus es 1.186.586. Todas las tesis de TeLing3N fueron publicadas en español entre los años 2000 a 2009. Además, todas ellas fueron escritas por un solo autor, pues, como hemos señalado, nos interesa la comunicación del conocimiento propio. Por otro lado, nos hemos asegurado de que TeLing3N cumpla con los requerimientos señalados por autores especialistas en trabajos de corpus, a saber: claridad sobre la procedencia de los textos e integridad (Parodi, 2010), reputación y accesibilidad (Nwogu, 1997).

Las tesis del corpus están etiquetadas manualmente con las FCCP que contienen. Para ello, TeLing3N se dividió en un corpus de prueba (33%), que fue la base para

la detección de las diferentes FCCP, y un corpus de investigación (67%), en el que se etiquetaron las FCCP detectadas en el corpus de prueba. Concretamente, para realizar la anotación de los textos, se siguieron los criterios siguientes:

- a. Identificación de la fuente o voz comunicante de la información, esto es, la fuente de origen de la información presentada en la tesis. Para ello, se realizó la distinción entre el contenido asociado lingüísticamente al tesista y a los autores citados. Para diferenciar ambos tipos de fuentes, se utilizó una serie de pistas lingüísticas.
- b. Identificación de la función concreta de la fuente propia: primero, se delimitó la noción de función como una categoría pragmática que refleja la intencionalidad del tesista en el texto, vale decir, la intención con que el autor de la tesis incorpora ciertos fragmentos en su trabajo (Meza, 2013). Para la delimitación de los fragmentos asociados a una función, seguimos las recomendaciones de Swales (2004), quien señala que, en un análisis de funciones del discurso, aunque a veces se ha alineado con una unidad gramatical, como una frase, una palabra, o un párrafo, es mejor considerarlo de manera flexible en cuanto a su realización lingüística. Así, una función puede ser realizada, en un caso, por una cláusula y, en otro, por varias frases.
- c. Registro de pistas lingüísticas: un paso que se realizó de forma paralela a los antes mencionados fue el registro de las pistas lingüísticas asociadas a cada una de las categorías levantadas. Esto nos permitió construir una base de datos con las marcas lingüísticas mediante las cuales se presentan, y gracias a las cuales pudimos reconocer más fácilmente cada FCCP.

Para determinar la consistencia de la anotación, se recurrió a cinco anotadores (todos doctores en lingüística y hablantes nativos de lengua española) y se midió el acuerdo entre ellos (*interannotator agreement*). Para ello se empleó el coeficiente Kappa de Fleiss (1971). El cálculo de este coeficiente indicó que el grado de acuerdo entre los evaluadores fue de 0,83 para el corpus de prueba y 0,73 para el de investigación, lo que, en términos de Landis y Koch (1977), corresponde a un acuerdo ‘casi perfecto’ en el primer caso y ‘considerable’ en el segundo.

En segundo lugar, a partir del corpus TeLing3N, constituimos un subcorpus para esta investigación conformado por 10 ejemplos de cada FCCP, seleccionados aleatoriamente. Para cada ejemplo se consideró la oración donde se muestra la FCCP, así como también las oraciones inmediatamente anterior y posterior (es decir, 3 oraciones por cada ejemplo). En caso de que no existiesen 10 ejemplos para una función, se incluyeron únicamente los ejemplos detectados. El subcorpus final consta de 386 ejemplos (1.158 oraciones).

En tercer lugar, cada ejemplo se anotó manualmente con las RD de la RST incluidas en la Tabla 2. Esta anotación fue realizada por una doctora experta en la RST con amplia experiencia en anotación discursiva de corpus. En la Tabla 3, para ejemplificar los resultados del análisis, presentamos dos casos reales. En el primero se muestra la FCCP ‘presentar la investigación o aspectos de ella’, que se corresponde con la RD de

‘propósito’. En el segundo, se presenta la FCCP ‘sintetizar un planteamiento teórico’, relacionada con la RD de ‘resumen’.

TABLA 3. EJEMPLOS DE CORRELACIÓN ENTRE FCCP Y RD

	Ejemplo 1	Ejemplo 2
FCCP	Presentar la investigación o aspectos de ella	Sintetizar un planteamiento teórico
RD	Propósito	Resumen
Oración del corpus	Esta investigación es un estudio preliminar, que aborda un tema de alto interés psicolingüístico —reconocimiento auditivo de palabras en <i>mapudungu</i> — e intenta hacerlo con rigor, tratando de superar las limitaciones derivadas de realizar un estudio experimental en una comunidad indígena con hablantes no urbanos.	En síntesis, determinando que un texto es un conjunto de proposiciones que se pueden agrupar en microproposiciones y macroproposiciones, la coherencia estaría relacionada con estas “estructuras”.

En cuarto lugar, se cuantificaron en términos de frecuencia absoluta (número de casos) y relativa (porcentajes) las RD de la RST detectadas en los 10 ejemplos del subcorpus analizado correspondientes a cada una de las FCCP de la lista incluida en la Tabla 1. En la Tabla 4 se muestra un ejemplo de la tabla empleada en nuestro trabajo para realizar esta cuantificación.

TABLA 4. EJEMPLO DE ANÁLISIS CUANTITATIVO QUE RELACIONA LA FCCP ‘PRESENTAR LA INVESTIGACIÓN O ASPECTOS DE ELLA’ CON LAS RD DE LA RST

Función	Nº de ejemplos	Relaciones RST detectadas	Frecuencia absoluta de relaciones	Frecuencia relativa de relaciones
Presentar la investigación o aspectos de ella	10	Propósito	9	90%
		Medio	1	10%

En quinto lugar, se observó si existían regularidades entre la aparición de las FCCP y las RD. Para ello, establecimos un umbral con el objetivo de determinar a partir de cuántos casos de coincidencia entre FCCP y RD se podía afirmar que existe relación relevante entre ellas. Específicamente, fijamos dicho umbral en un 60%, número que ha sido considerado por otras investigaciones de análisis de género para establecer el estatus de una categoría, en las que se señala que es el umbral adecuado para establecer el estatus de una categoría en estudios relacionados con géneros textuales (Ibáñez, 2008;

Kanoksilapatham, 2005, 2007; Martínez 2012). Concretamente, Kanoksilapatham (2005, 2007) utiliza dicho umbral para establecer el estatus obligatorio u opcional de una movida retórica en AIC. Ibáñez (2008) lo utiliza para distinguir entre los propósitos comunicativos del género Texto Disciplinar y clasificarlos en nucleares u opcionales. Por su parte, Martínez (2012) lo considera para determinar la variación en la organización retórico-funcional en el género tesis, estableciéndolo como punto de corte para definir el estatus de las movidas retóricas identificadas.

Por ejemplo, en la Tabla 5, se observa que, en los 10 ejemplos de la función ‘indicar vacíos/necesidades en el conocimiento’, el 30% incluye una RD de ‘antítesis’, el 60% una RD de ‘interpretación’ y el 10% una RD de ‘propósito’. Así, la única correspondencia que consideramos relevante es con la relación de ‘interpretación’.

TABLA 5. EJEMPLO DE ANÁLISIS CUANTITATIVO QUE RELACIONA LA FCCP ‘INDICAR VACÍOS/NECESIDADES EN EL CONOCIMIENTO’ CON RD DE LA RST

Función	Nº de ejemplos	Relaciones RST detectadas	Frecuencia absoluta de relaciones	Frecuencia relativa de relaciones
Indicar vacíos/ necesidades en el conocimiento	10	Antítesis	3	30%
		Interpretación	6	60%
		Propósito	1	10%

Asimismo, en nuestro estudio no consideramos la relación discursiva de ‘elaboración’, ya que en el marco de la RST es la relación más genérica y no suele incluir marcas explícitas que la evidencien. Esto es debido a que la función de un satélite de este tipo es, simplemente, ofrecer información adicional de tipo general.

Una vez realizado el análisis cuantitativo del corpus, se extrajo un listado de las correspondencias entre FCCP y RD que superaban el 60% de frecuencia relativa. Finalmente, a partir de los resultados obtenidos, se proponen estrategias de ayuda para estudiantes que deban elaborar una tesis.

5. Análisis y resultados de la correspondencia entre las FCCP y las RD de la RST

En este apartado ofrecemos los resultados del análisis cuantitativo realizado sobre el corpus junto con un análisis cualitativo de los casos más relevantes que hemos detectado. La lista completa de las correspondencias entre las FCCP y las RD en el corpus de análisis puede observarse en el Anexo 1. A partir de esta lista de correspondencias, se extrajeron las FCCP que tienen una relación relevante con una RD (más del 60%), incluidas en la Tabla 6.

Como se ve en la Tabla 6, de las 40 FCCP iniciales, solo 25 (62,5%) tienen una relación relevante entre la FCCP y la RD. Nos centraremos en ellas para ofrecer a los

estudiantes las marcas empleadas con mayor frecuencia para expresar las FCCP y, así, ayudarles a redactar sus tesis. De estas 25 FCCP, 12 (30%) se corresponden en un 100% con una única RD, lo que evidencia una correlación muy directa entre ellas. Es el caso, por ejemplo, de la función ‘problematizar’ con la RD ‘preparación’ y de ‘indicar proyecciones’ con la RD ‘interpretación’, entre otros casos. En cambio, los 13 casos restantes (32,5%) tienen una correlación menor del 100% pero superior al 60%.

TABLA 6. LISTA DE CORRESPONDENCIAS RELEVANTES ENTRE FCCP Y RD EN EL CORPUS

Función	Tipo de relaciones RST	%
Presentar la investigación o aspectos de ella	Propósito	90%
Indicar vacíos/necesidades en el conocimiento	Interpretación	60%
Problematizar	Preparación	100%
Justificar la investigación	Justificación	90%
Indicar organización de la investigación	Secuencia	80%
Aludir a la tesis misma anafóricamente	Reformulación	90%
Indicar proyecciones	Interpretación	100%
Indicar objetivos	Propósito	100%
Describir una metodología o instrumento propio	Medio	100%
Explicar procedimientos	Medio	100%
Definir un concepto	Fondo	100%
Enmarcar la investigación o un concepto relevante para ella	Fondo	100%
Explicar un concepto o teoría	Interpretación	70%
Emitir juicios sobre un concepto o teoría	Interpretación	80%
Indicar acuerdo con un autor y/o su planteamiento	Interpretación	100%
Sintetizar un planteamiento teórico	Resumen	100%
Aplicar teorías o investigaciones a otros contextos	Interpretación	90%
Proponer aspectos teóricos	Fondo	60%
Indicar opción teórica	Medio	70%
Introducir conceptos	Fondo	100%
Exponer hallazgos de la investigación	Resultados	60%
Describir material didáctico	Medio	100%
Realizar interpretaciones o deducciones	Interpretación	90%
Contradecir investigaciones previas	Antítesis	62,5%
Refutar hipótesis de investigación	Interpretación	100%

En cuanto al análisis cualitativo de los datos obtenidos, destacamos varios aspectos. En primer lugar, de las 40 FCCP analizadas, 26 de ellas (65%) se correlacionan con más de una RD. Es el caso, por ejemplo, de la FCCP ‘corroborar investigaciones previas’, que se correlaciona con las RD ‘elaboración’ (20%), ‘propósito’ (10%), ‘lista’ (10%), ‘resultados’ (10%), ‘interpretación’ (40%) y ‘antítesis’ (10%). Esto quiere decir que la manera de expresar esta FCCP en el género tesis no es sistemática y puede reflejarse en el texto a través de diversas estrategias discursivas.

En segundo lugar, encontramos cinco casos en los que una FCCP se correlaciona con una RD específica, pero en el interior de la oración existe otra RD adicional diferente. Por ejemplo, la FCCP ‘destacar la importancia de un concepto o teoría’ se correlaciona en el 100% de los casos con la RD de ‘interpretación’. Sin embargo, en un 60% de ellos, en el interior de la oración de la FCCP se incluye otra RD. Concretamente, los patrones encontrados para esta FCCP son: ‘interpretación’ (+‘causa’) (20%), ‘interpretación’ (+‘concesión’) (20%), ‘interpretación’ (+‘antítesis’) (10%) e ‘interpretación’ (+‘reformulación’) (10%). La Figura 2 muestra un ejemplo. En este árbol discursivo, el conjunto 2-3 equivale a la oración que refleja la FCCP y constituye un satélite de la ‘interpretación’ del segmento anterior (1). A su vez, dicho conjunto incluye dos segmentos discursivos: el segmento 3 es un satélite de ‘causa’ del segmento 2, que es su núcleo.

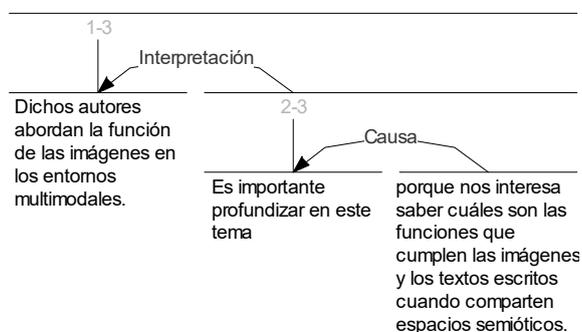


FIGURA 2. EJEMPLO DE ÁRBOL DISCURSIVO DE LA RST QUE MUESTRA UNA FCCP QUE INCLUYE DOS RD.

Estos casos evidencian que hay FCCP más complicadas de expresar que otras. Por tanto, en estos casos más complejos se necesitan varias RD; por ejemplo, para argumentar la importancia de un concepto o una teoría, como ocurre en este caso particular.

En tercer lugar, hemos observado seis FCCP que tienen asignada una relación de ‘elaboración’ en más del 60% de los casos. De estas, tres tienen incluso una correlación del 100%: ‘explicar un gráfico, cuadro o figura de otro’, ‘explicar un gráfico, cuadro o figura propio’ y ‘ejemplificar un concepto o teoría’. Estos seis casos superan el umbral fijado en esta investigación para establecer una relación relevante entre una FCCP y una RD. Sin embargo, no los tenemos en cuenta porque, como ya se ha mencionado,

la RD de ‘elaboración’ es la más frecuente y general en la lengua, y se puede expresar de múltiples maneras. Por tanto, nos aportará poca información útil para el diseño de estrategias de ayuda a la redacción.

6. Estrategias para redacción de FCCP en el género tesis

A partir de los resultados obtenidos y de la confirmación de que sí existe una correlación sustancial entre determinadas FCCP y RD, proponemos dos estrategias complementarias para los estudiantes que deban redactar tesis y, por tanto, incluir en ellas FCCP. Por un lado, les remitimos al *RST Spanish Treebank* para extraer ejemplos obtenidos de textos reales que muestren RD que se correspondan con las FCCP que desean expresar. Por otro, ofrecemos el conjunto de marcas lingüísticas que se utilizan en nuestro corpus para evidenciar las FCCP.

6.1. Consulta del RST Spanish Treebank

El *RST Spanish Treebank* es un corpus que incluye 267 textos especializados en español que se dividen en nueve ámbitos especializados y se corresponden con diversos géneros textuales. Todos los textos están anotados manualmente con estructuras discursivas de la RST mediante la RSTTool. El 69% (183 textos) fue anotado por un único anotador y el 31% (84 textos) por dos anotadores. La interfaz de consulta del corpus está en línea de manera gratuita, tal como otras recientes interfaces para buscar automáticamente relaciones discursivas en corpus (Iruskieta et al., 2013). Desde dicha interfaz se puede descargar el corpus anotado, visualizar los diferentes textos anotados y también utilizar tres herramientas que permiten extraer diferentes tipos de información del corpus. Una de estas herramientas permite al usuario extraer un listado de los fragmentos del corpus que incluyen una RD específica. Para ello, el usuario simplemente debe seleccionar la herramienta RST_extract (v. Figura 3) y seleccionar una de las RD de la lista que se ofrece (v. Figura 4).

A continuación, el sistema ofrecerá al usuario el listado de casos del corpus que se corresponden con satélites de dicha RD, en caso de tratarse de una RD núcleo-satélite. Por ejemplo, si un estudiante quisiera redactar la FCCP ‘aludir a la tesis misma anafóricamente’, debería seleccionar la RD de ‘reformulación’ para obtener ejemplos representativos en los que basarse, ya que, según nuestro análisis y tal como muestra la Tabla 6, existe una correlación del 90% entre ambas. En este caso, se incluyen en la Figura 5 algunos de los ejemplos obtenidos por la interfaz, en los que hemos destacado en negrita los marcadores que evidencian la RD de ‘reformulación’.

En caso de querer obtener ejemplos representativos para la redacción de la FCCP ‘sintetizar un planteamiento teórico’, el estudiante debería seleccionar la RD de ‘resumen’ (existe una correlación del 100% entre ambas en nuestro corpus). En este caso, obtendría, por ejemplo, los casos incluidos en la Figura 6, en los que hemos destacado en negrita los marcadores que evidencian la RD de ‘resumen’.

The screenshot displays the RST Spanish Treebank interface. At the top, there is a navigation bar with buttons for INICIO, CORPUS, MANUAL RST, PROYECTO, CONTACTO, and FAQ. Below this, the 'Procesar' section includes buttons for 'Guardar selección', 'Cargar selección', 'Descargar Corpus', and 'Procesar'. A dropdown menu is open, showing options: RST_stats_rel, RST_extract, and RST_stats_EDUs. The 'CORPUS' section is expanded, showing a tree structure of domains: Astrofísica, Derecho, Economía, Ingeniería sísmica, Lingüística, Adquisición del lenguaje, Lingüística aplicada, and PLN. Under 'PLN', several sub-domains are listed, including Terminología, Matemáticas, Medicina, Psicología, and Sexualidad. The 'Información' section on the right provides details for a selected document: 'Inducción de un Lexicón de Opinión Orientado al Dominio', ID: ii00045, Palabras: 136, and a link to the SEPLN website. It also lists the authors (Cruz, F.; Troyano, J.A.; Ortega, J.; Vallejo, C.G.) and the journal (Procesamiento de Lenguaje Natural, 43, 5-12). The 'Texto' section shows the document's content with semantic annotations (e.g., 'Medio', 'Evaluación', 'Lista', 'Secuencia', 'Resultado') and their associated scores. The 'Archivo seleccionado' section shows a detailed view of the document's content with semantic annotations and scores.

FIGURA 3. CAPTURA DE PANTALLA DE LA INTERFAZ DE CONSULTA DEL *RST SPANISH TREEBANK*.

This close-up screenshot shows a dropdown menu from the RST Spanish Treebank interface. The menu is open, displaying a list of options: Elaboración, Evidencia, Preparación, Fondo, Resultado, and Justificación. The menu is positioned over a text input field and a 'Procesar' button.

FIGURA 4. DETALLE DE LA INTERFAZ DE CONSULTA DEL *RST SPANISH TREEBANK*.

esto es, se tomaron en cuenta los rasgos específicos de esta lengua y, en particular, la posibilidad de que en una cláusula del español el verbo aparezca en posición inicial.

esto es, el grupo de menor reconocimiento social, las mujeres, no perciben mayor discriminación que los varones.

es decir, hemos de reducir las características de los mismos.

es decir, entender los mecanismos de transmisión de la política monetaria.

es decir, que pueden utilizar la alternancia de códigos para dos cosas:

(**es decir**, cuánto hay que girar por el eje vertical)

en otras palabras, el asunto consiste en resolver cuáles fuentes de prueba pueden ser incorporadas a un juicio como medios de prueba relevantes y jurídicamente admisibles.

FIGURA 5. EJEMPLOS DE ‘REFORMULACIÓN’ OBTENIDOS POR LA HERRAMIENTA RST_EXTRACT.

En resumen, el objetivo del presente trabajo consiste en conseguir una sistematización que nos ayudaría a encontrar de un modo bastante automático el equivalente en euskera de adjetivos referenciales de las lenguas vecinas.

En resumen, la clasificación de las relaciones que puede darse entre los nombres y sus modificadores es universal: por una parte los modificadores con carácter de predicado y por otra los que tienen carácter referencial.

Se presenta una síntesis de los principales enfoques teóricos y resultados de investigación sobre el desarrollo lingüístico, con especial atención a las bases biológicas y neuroanatómicas de la adquisición del lenguaje.

FIGURA 6. EJEMPLOS DE ‘RESUMEN’ OBTENIDOS POR LA HERRAMIENTA RST_EXTRACT.

En el caso de las relaciones multinucleares, la herramienta RST_extract ofrece al usuario la lista de los núcleos relacionados que se corresponden con la RD seleccionada. Por ejemplo, si el estudiante quiere obtener ejemplos de fragmentos que reflejen la FCCP ‘indicar organización de la investigación’ debería seleccionar la RD multinuclear

de ‘secuencia’ (ya que en nuestro corpus existe una correlación entre ambas del 80%). Así, obtendrá, por ejemplo, los casos mostrados en la Figura 7 (las dos barras diagonales indican frontera de segmento discursivo, es decir, núcleos diferentes), donde hemos destacado en negrita los marcadores organizativos usados para reflejar la RD de ‘secuencia’.

// Con este propósito, **en la primera parte** se revisan los elementos centrales en los que éste se sustenta.// **En la segunda parte** del artículo se procede a evaluar críticamente tales elementos.//

// Este artículo analiza, **en primer lugar**, el impacto de la Regeneración (1880-1899) en el sistema monetario y bancario colombiano [...]// **En segundo lugar**, estudia el difícil proceso de recuperación económica y el reordenamiento monetario al inicio del siglo XX.//

// **primera** porque el habla es uno de los indicadores del desarrollo integral del niño,// **segunda** para detectar las dificultades en este proceso// y **tercera** para determinar los niños en riesgo.//

// Léelo con cuidado// y, **a continuación**, completa las tablas con los datos que aparecen en la noticia.//

// **En la primera parte** se proporcionarán rápidamente las definiciones pertinentes: esfera unitaria, contraíble y espacio.// El lector familiarizado con todos los términos arriba mencionados puede pasar sin remordimiento a **la segunda parte**, en donde se verá un poquito de historia relacionada con el tema y qué tiene que ver todo esto con el Teorema del Punto Fijo.// **En la tercera y última sección** se encuentra la demostración prometida.//

// **A continuación** se presenta una reseña biográfica de Riemann.// **Posteriormente** una reseña de la tesis de Riemann.// **Finalmente** se presenta la traducción de la tesis.//

// **En la primera parte** se describe el análisis de la interacción de interés,// **en la segunda** la evaluación de lo referido por el usuario y las condiciones encontradas,// **la tercera parte** detalla la regulación,// **para finalmente en la cuarta** mostrar parte del seguimiento y la probabilización de la conducta.//

FIGURA 7. EJEMPLOS DE ‘SECUENCIA’ OBTENIDOS POR LA HERRAMIENTA RST_EXTRACT.

En los tres casos anteriores, cada una de las RD se corresponde con una única FCCP. Sin embargo, también encontramos casos en los que una RD está correlacionada con diversas FCCP, como ‘interpretación’, que se correlaciona en diferentes porcentajes (superiores al 60%) con ocho FCCP. Si seleccionamos esta RD mediante la herramienta RST_extract, se obtendrán fragmentos como los mostrados en la Figura 8, en los que hemos destacado en negrita los marcadores (principalmente estructuras sintácticas) que ayudan a evidenciar la RS de ‘interpretación’.

por lo que se dice que no hay motores internos para activar la economía, y se está dependiendo primordialmente de factores foráneos [...].

De esta manera, se puede inferir que la política monetaria ha mostrado una evolución favorable en términos de su impacto sobre las tasas de interés de mercado, desde el anuncio del corredor de referencia y la adopción del esquema MEI.

lo que pone de manifiesto una serie de presiones internas, aun a favor del mantenimiento de la lengua indígena.

y, por tanto, la política monetaria chilena **puede decirse** que era única en el mundo en cuanto a utilizar como objetivo intermedio una tasa real de interés.

Es probable que ellos disminuyan paulatinamente a medida que los individuos pasen a concebir los valores reales en términos del nivel de precios esperado.

El peligro de que esto ocurra es que se rompa el equilibrio necesario entre un cierto grado de arbitrariedad, inevitable para el consenso, y el uso real de la terminología por parte de los especialistas.

Nuestra hipótesis es que una característica sintáctica del euskera y de las lenguas románicas se extrapola hasta la morfología [...].

Esto nos lleva a una cuestión fundamental: la terminología de Internet traspasa los límites del área de especialidad (a la que se circunscribe por definición el léxico científico y técnico) e irrumpe en la lengua de uso general [...].

lo que supone una aportación importante a la lexicografía actual [...].

FIGURA 8. EJEMPLOS DE ‘INTEPRETACIÓN’ OBTENIDOS POR LA HERRAMIENTA RST_EXTRACT.

6.2 Marcas lingüísticas utilizadas para expresar FCCP

El uso del conjunto de marcas lingüísticas con las que cada FCCP se presenta en las tesis incluidas en nuestro corpus es el segundo tipo de estrategia que proponemos para ayudar a los estudiantes en la escritura de su tesis. Específicamente, corresponden a pistas de distintos tipos (marcadores o conectores, verbos, formas verbales, unidades léxicas, etc.) que hemos extraído de nuestro corpus anotado.

Estas pistas suponen un recurso de gran utilidad para quienes deban escribir una tesis, pues son ejemplos reales de las marcas lingüísticas utilizadas para dar cuenta de las distintas FCCP que se correlacionan sustancialmente con RD. Estas marcas han sido extraídas de manera literal del corpus analizado. Todas las marcas de las FCCP que se correlacionan positivamente con una RD se encuentran disponibles para su consulta en línea¹.

Por ejemplo, si un estudiante requiere utilizar en su tesis las FCCP ‘contradecir investigaciones previas’, ‘exponer hallazgos de la investigación’ o ‘introducir conceptos’, puede revisar el archivo disponible en dicho vínculo y obtener las pistas lingüísticas precisas con las que podría redactar dichas FCCP, tal como se muestra en la Tabla 7. Así, ofrecemos al estudiante la información específica que necesita para cumplir con un contenido obligatorio del género tesis, que es expresar sus propias aportaciones.

TABLA 7. EJEMPLOS DE MARCAS LINGÜÍSTICAS PARA LAS FCCP ‘CONTRADECIR INVESTIGACIONES PREVIAS’, ‘EXPONER HALLAZGOS DE LA INVESTIGACIÓN’ E ‘INTRODUCIR CONCEPTOS’

Función	Tipo de relación RST	Tipos de marcas lingüísticas de la función	Marcas lingüísticas de la función
Contradecir investigaciones previas	Antítesis	Formas verbales	refuta- contraponen- difiere- (no) concordamos
		Marcadores discursivos	por el contrario- sin embargo- si bien- en cambio- pero- a pesar de- pero
		Palabras/ frases clave	en contraste con los estudios anteriormente citados, nuestros datos refutan X hipótesis o al menos plantean dudas al respecto- una prueba en contra de la hipótesis de X autor es...- se contraponen taxativamente- difiere de- este resultado se contrapone con- contrario a lo que indican X estudios- a diferencia de X autor/investigación- no concordamos con X autor- contrario a lo que podría creerse...- nuestros datos desmienten X hipótesis/ autor/ investigación

1. <https://recursoslinguisticosblog.wordpress.com/marcas-linguisticas/expresion-del-conocimiento-propio-en-tesis/>

Exponer hallazgos de la investigación	Resultados	Formas verbales	se descubrió- encontramos- hemos obtenido- indica- ejemplifica- identificado- hemos observado
		Palabras/ frases clave	de acuerdo a los datos que hemos obtenido...- nos encontramos con algunos interesantes resultados, tales como...- de acuerdo a nuestros datos...- con respecto a los resultados...- los resultados de la comparación indican X- las diferencias no son significativas- este resultado indica X- estos gráficos ejemplifican X- según los resultados del análisis...- en nuestros datos encontramos X- este rasgo ha sido identificado en X% de los sujetos- en nuestro análisis encontramos X- lo más/ menos frecuente es X- X categoría tiene la mayor/menor frecuencia de aparición
Introducir conceptos	Fondo	Formas verbales	entendemos- definimos- proponemos- se concibe- llamaremos- denominamos
		Palabras/ frases clave	para los propósitos de nuestra investigación, se ha definido X noción como...- para nosotros serán útiles las nociones X e Y- entendemos como- se concibe como- proponemos una noción teórica nueva- X fenómeno corresponde a lo que podemos llamar como Y- no nos permitimos llamar X a estas expresiones, sino que las denominamos Y- lo que denominamos X se define como...

Como se observa en la Tabla 7, la FCCP ‘contradecir investigaciones previas’ se expresa, principalmente, acudiendo a tres tipos de marcas lingüísticas: formas verbales, marcadores discursivos, y ciertas palabras o frases clave. Por su parte, las FCCP ‘exponer hallazgos de la investigación’ e ‘introducir conceptos’ se expresan sobre todo mediante marcadores discursivos, y palabras o frases clave.

7. Conclusiones y trabajo futuro

Como se ha explicado en este trabajo, una de las principales dificultades a las que se enfrentan los estudiantes que deben redactar una tesis es expresar el conocimiento propio adecuadamente. No obstante, existen pocos trabajos que aborden esta cuestión. En este trabajo partimos de la hipótesis de que existen algunas FCCP en el género tesis que tienen una correspondencia relevante con las RD de la RST. El objetivo teórico-descriptivo de la investigación fue determinar entre qué FCCP y RD existe una correspondencia

sustancial en dicho género. El objetivo aplicado fue proponer estrategias discursivas para la redacción de las FCCP en el género tesis.

Los resultados cuantitativos del estudio, basado en corpus, indican que sí existen correspondencias relevantes entre algunas FCCP y RD, mientras que hay otras que no presentan dicha correspondencia o no es relevante. Concretamente, de las 40 FCCP de las que se partió en el trabajo, 25 (62,5%) tienen una relación relevante entre la FCCP y la RD. Por tanto, la hipótesis inicial se ha validado y el objetivo teórico-descriptivo ha sido logrado.

El establecimiento de esta correlación nos llevó al objetivo aplicado del trabajo y nos permitió diseñar dos estrategias discursivas complementarias para ayudar a los estudiantes a expresar el conocimiento propio en sus tesis: en primer lugar, una estrategia de extracción de ejemplos representativos del *RST Spanish Treebank* y, en segundo lugar, una estrategia basada en un listado de pistas lingüísticas extraído de manera literal de nuestro corpus.

Como trabajo futuro nos planteamos buscar anotadores adicionales para anotar una parte del corpus y posteriormente medir el *interannotator agreement* entre ellos a través del coeficiente Kappa de Fleiss (1971), con el objetivo de corroborar la estabilidad de la anotación. Asimismo, nos gustaría aplicar técnicas estadísticas avanzadas como complemento al criterio de frecuencia basado en porcentajes empleado en este estudio, con el propósito de obtener información sobre la significación estadística de los resultados. También tenemos prevista una evaluación de nuestra propuesta a partir de su uso por parte de estudiantes universitarios.

Finalmente, a partir de los resultados obtenidos en este estudio, planeamos el desarrollo de un sistema semiautomático de asistencia para la redacción de contenidos en el género tesis, dirigido a estudiantes de licenciatura, máster y doctorado. El sistema incorporará estrategias discursivas y pistas lingüísticas que ayuden al estudiante a redactar las FCCP de su tesis que están correlacionadas sustancialmente con RD de la RST. Actualmente ya existen sistemas automáticos de ayuda a la redacción de textos que van en esta línea, como se comentó en el apartado 2.3.

Agradecimientos

El trabajo ha sido parcialmente financiado gracias al apoyo de la Dirección de Investigación de la Universidad de La Serena, Chile. Este trabajo también ha sido parcialmente financiado por el contrato Ramón y Cajal (RYC-2014-16935) de Iria da Cunha en el Departamento de Filologías Extranjeras y sus Lingüísticas de la Facultad de Filología de la Universidad Nacional de Educación a Distancia en España. Asimismo, se enmarca en los grupos de investigación ACTUALing e IULATERM.

Referencias

- Acrolinx GmbH (s. f.). *Acrolinx* [Software]. Recuperado de <http://www.acrolinx.com/>
arText = da Cunha (2016)
- Botta, M. (2002). *Tesis, monografías e informes: Nuevas normas y técnicas de investigación y redacción*. Buenos Aires: Biblos.
- Bui, Y. (2013). *How to write a master's thesis*. Los Ángeles: Sage.
- CALeSE = Núcleo Interinstitucional de Lingüística Computacional (s. f.)
- Carlson, L., Marcu, D. y Okurowski, M. E. (2001). Building a discourse-tagged corpus in the framework of Rhetorical Structure Theory. En *Proceedings of the 2nd SIGDIAL Workshop on Discourse and Dialogue* (pp. 1-10). Stroudsburg: Association for Computational Linguistics.
- Cassany, D. (2007). *Afilar el lapicero. Guía de redacción para profesionales*. Barcelona: Anagrama.
- da Cunha, I. (2016). *El trabajo de fin de grado y de máster: Redacción, defensa y publicación*. Barcelona: Editorial UOC.
- da Cunha, I. (Dir.) (2016). *Sistema arText* [Software]. Recuperado de <http://sistema-artext.com/>
- da Cunha, I., Torres-Moreno, J. M., y Sierra, G. (2011a). On the Development of the RST Spanish Treebank. En N. Ide, A. Meyers, S. Pradhan y K. Tomarek, *Proceedings of the 5th Linguistic Annotation Workshop. 49th Annual Meeting of the Association for Computational Linguistics (ACL)* (pp. 1-10). Stroudsburg: Association for Computational Linguistics.
- da Cunha, I., Torres-Moreno, J. M., y Sierra, G. (2011b). *RST Spanish Treebank* [Software]. Disponible en http://corpus.iingen.unam.mx/rst/index_es.html
- da Cunha, I., SanJuan, E.; Torres-Moreno, J. M., Cabré, M. T., y Sierra, G. (2012). A Symbolic approach for automatic detection of nuclearity and rhetorical relations among intra-sentence discourse segments in spanish. En a. Gelbukh (ed.), *Computational linguistics and intelligent text processing. CICLing 2012. Lecture notes in Computer Science*, vol. 7181, pp. 462-474. Berlín: Springer.
- Eckel, E. (2011). Textual appropriation in engineering Master's theses: A preliminary study. *Science and engineering ethics*, 17(3), 469-483.
- Eco, U. (2006). *Cómo se hace una tesis: Técnicas y procedimientos de estudios, investigación y escritura*. Barcelona: Gedisa.
- eWritingPal (s. f.). *eWritingPal* [Software]. Recuperado de <http://www.ewritingpal.com/>
- Fleiss, J. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76(5), 378-382.
- GrammarChecker = Lawley et al. (2015)
- Grammarly = Lytvyn et al. 2009

- Harwood, N. (2005). 'I hoped to counteract the memory problem, but I made no impact whatsoever': Discussing methods in computing science using I. *English for Specific Purposes*, 24(3), 243-267.
- Hidalgo, C., y Passarella, V. (2009). Encuesta 'Escritura y producción de conocimiento en carreras de Posgrado' administrada a estudiantes de posgrado de diversas instituciones de Argentina y Chile (2005-2006). En E. Arnoux (Ed.), *Escritura y producción de conocimiento en las carreras de posgrado* (pp. 62-73). Buenos Aires: Santiago Arcos Editor.
- Ibáñez, R. (2008). El texto disciplinar y el acceso al conocimiento desde el análisis del género: ¿Regulación del conocimiento o persuasión? En G. Parodi (Ed.), *Géneros académicos y géneros profesionales: accesos discursivos para saber y hacer* (pp. 219-246). Valparaíso: Ediciones Universitarias de Valparaíso.
- Iruskieta, M., Aranzabe, M., Díaz de Ilaraza, A., González, I., Lersundi, I. y López de Lacalle, O. (2013). The RST Basque TreeBank: An online search interface to check rhetorical relations. En *Anais do IV Workshop "A RST e os Estudos do Texto"* (pp. 40-49). Fortaleza: Sociedade Brasileira de Computação.
- Iruskieta, M., Díaz de Ilaraza, A. y Lersundi, M. (2015). Establishing criteria for RST-based discourse segmentation and annotation for texts in Basque. *Corpus Linguistics and Linguistic Theory*, 11(2), 303-334.
- Kanoskilapatham, B. (2005). Rhetorical structure of biochemistry research articles. *English for Specific Purposes*, 24(3), 269-292.
- Kanoskilapatham, B. (2007). Rhetorical moves in biochemistry research articles. En D. Biber, U. Connor y T. Upton (Eds.), *Discourse on the move* (pp. 73-119). Amsterdam: Benjamins.
- Kuo, C. (1999). The use of personal pronouns: Role relationships in scientific journal articles. *English for Specific Purposes*, 18(2), 121-138.
- Landis, J. y Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159-174.
- LanguageTool = Rysin et al. (s. f.)
- Lawley, L., Chacón, R. y Martín, S. (2015). *GrammarChecker* [Software]. Recuperado de https://www.e-uned.es/correctme/cm_english/
- Lytvyn, M., Shevchenko, A. y Dmytro L. (2009). *Grammarly* [Software]. Recuperado de <http://www.grammarly.com/>
- Mann, W. C. y Taboada, M. (2005). *RST Web Site*. Recuperado de www.sfu.ca/rst
- Mann, W. C. y Thompson, S. A. (1988). Rhetorical Structure Theory: Toward a functional theory of text organization. *Text*, 8(3), 243-281.
- Martínez, J. (2012). *Descripción y variación retórico-funcional del género tesis doctoral: Un análisis desde dos disciplinas y dos comunidades discursivas a partir del Corpus Te DiCE-2010* (Tesis doctoral). Pontificia Universidad Católica de Valparaíso.
- Martínez, J. (2015). El género tesis doctoral de historia y física: Descripción y variación retórico-funcional. En G. Parodi y G. Burdiles (Eds.), *Leer y escribir en contextos*

- académicos y profesionales: Géneros, corpus y métodos* (pp. 113-152). Valparaíso: Ediciones Universitarias de Valparaíso.
- Maulid, W. (2017). The moves and steps in the literature review and discussion sections in the four master's degree theses. *Journal of English Education*, 2(1), 51-58.
- MeaningCloud (s. f.). *Stilus* [Software]. Recuperado de <http://www.mystilus.com/>
- Meza, P. (2013). *La comunicación del conocimiento en las secciones de tesis de lingüística: determinación de la variación entre grados académicos* (Tesis doctoral). Pontificia Universidad Católica de Valparaíso.
- Meza, P. (2015). La comunicación del conocimiento en el género tesis de lingüística: Comparación entre grados académicos. En G. Parodi y G. Burdiles (Eds.), *Leer y escribir en contextos académicos y profesionales: Géneros, corpus y métodos* (pp. 67-112). Valparaíso: Ediciones Universitarias de Valparaíso.
- Meza, P y Sabaj, O. (2016). Funciones discursivas de consenso y disenso en tesis de lingüística. *Onomázein*, 33, 385-411.
- Montolío, E. (2014). *Manual de escritura académica y profesional*. Barcelona: Ariel.
- Núcleo Interinstitucional de Lingüística Computacional (s. f.). *CALeSE* [Software]. Recuperado de <http://www.nilc.icmc.usp.br/calese/>
- Nwogu, K. (1997). The medical research paper: Structure and functions. *English for Specific Purposes*, 16(2), 119-138.
- O'Donnell, M. (2000). RSTTool 2.4: A markup tool for Rhetorical Structure Theory. En *Proceedings of the First International Conference on Natural Language Generation* (pp. 253-256). Stroudsburg: Association for Computational Linguistics.
- Paltridge, B., y Starfield, S. (2007). *Thesis and dissertation writing in a second language: A handbook for supervisors*. Londres: Routledge.
- Pardo, T., y Nunes, M. (2008). On the development and evaluation of a brazilian portuguese discourse parser. *Journal of Theoretical and Applied Computing*, 15(2), 43-64.
- Parodi, G. (2010). *Lingüística de corpus: De la teoría a la empiria*. Madrid: Iberoamericana Vervuert.
- RST Spanish Treebank* = da Cunha et al. (2011)
- Rysin, A., Blum, C., Naber, D., Pellé, D., Byagowi, E., Venugopal, E., ... y Reztsov, Y. (s. f.). *LanguageTool* [Software]. Recuperado de <https://www.languagetool.org/>
- Sabino, C. (2003). *Cómo hacer una tesis y elaborar todo tipo de escritos*. Bogotá: Panamericana.
- Samraj, B. (2008). A discourse analysis of master's theses across disciplines with a focus on introductions. *Journal of English for Academic Purposes*, 7(1), 55-67.
- Savio, A. (2010). Las huellas del autor en el discurso académico: Un estudio sobre tesis de psicoanalistas argentinos. *Lenguaje*, 38(2), 563-590.
- Stilus = MeaningCloud (s. f.)
- SWAN = University of Eastern Finland (s. f.)

- Taboada, M. (2016). Discourse markers as signals (or not) of rhetorical relations. *Journal of Pragmatics*, 38, 567-592.
- Tang, R., y John, S. (1999). The 'I' in identity: Exploring writer identity in student academic writing through the first person pronoun. *English for Specific Purposes*, 18, 23-39.
- Tapia, M. y Burdiles, G. (2012). La organización retórica del marco referencial en tesis de Trabajo Social. *Alpha (Osorno)*, 35, 169-184.
- Tofiloski, M., Brooke, J., y Taboada, M. (2009). A Syntactic and Lexical-Based Discourse Segmenter. En K. Y. Su, J. Su, J. Wiebe y H. Li (Eds.), *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers* (pp. 77-80). Stroudsburg: Association for Computational Linguistics.
- Swales, J. (2004). *Research genres: Explorations and applications*. Cambridge: Cambridge University Press.
- University of Eastern Finland (s. f.). *SWAN* [Software]. Recuperado de <https://cs.joensuu.fi/swan/>
- Venegas, R., Núñez, M. T., Zamora, S., y Santana, A. (2015). *Escribir desde la pedagogía del género. Guías para escribir el trabajo final de grado en licenciatura*. Valparaíso: Ediciones Universitarias de Valparaíso.
- Wuttisrisiriporn, N. (2017). Comparative rhetorical organization of ELT thesis introductions composed by Thai and American students. *English Language Teaching*, 10(12), 1-14.

Apéndice: Lista de correspondencias entre FCCP y RD en el corpus

Función	Nº de ejemplos	Nº de relaciones RST	Tipo de relaciones RST	%
1. Presentar la investigación o aspectos de ella	10	9	Propósito	90%
		1	Medio	10%
2. Indicar vacíos/necesidades en el conocimiento	10	3	Antítesis	30%
		6	Interpretación	60%
		1	Propósito	10%
3. Problematizar	10	10	Preparación	100%
4. Justificar la investigación	10	9	Justificación	90%
		1	Antítesis	10%
5. Indicar organización de la investigación	10	8	Secuencia	80%
		1	Circunstancia	10%
		1	Preparación	10%
6. Aludir a la tesis misma anafóricamente	10	9	Reformulación	90%
		1	Secuencia	10%
7. Comparar lo propio con lo de otros	10	3	Resultado	30%
		1	Justificación	10%
		3	Antítesis	30%
		3	Interpretación	30%
8. Explicar un gráfico, cuadro o figura de otro	10	10	Elaboración	100%
9. Explicar un gráfico, cuadro o figura propio	10	10	Elaboración	100%
10. Indicar dificultades	10	6	Elaboración	60%
		1	Antítesis	10%
		1	Concesión	10%
		2	Interpretación	20%
11. Indicar proyecciones	10	10	Interpretación	100%

Función	Nº de ejemplos	Nº de relaciones RST	Tipo de relaciones RST	%
12. Indicar limitación	10	5	Interpretación	50%
		2	Antítesis	20%
		1	Reformulación	10%
		1	Propósito (+ Concesión)	10%
		1	Elaboración	10%
13. Superar limitación	7	3	Solución	42,86%
		2	Antítesis	28,57%
		1	Reformulación	14,29%
		1	Lista	14,29%
14. Indicar hipótesis	10	1	Interpretación	10%
		1	Antítesis	10%
		5	Núcleo	50%
		2	Elaboración	20%
		1	Resultado	10%
15. Indicar objetivos	10	10	Propósito	100%
16. Describir una metodología o instrumento propio	10	10	Medio	100%
17. Explicar procedimientos	10	10	Medio	100%
18. Definir un concepto	9	9	Fondo	100%
19. Ejemplificar un concepto o teoría	10	10	Elaboración	100%
20. Enmarcar la investigación o un concepto relevante para ella	10	10	Fondo	100%
21. Explicar un concepto o teoría	10	3	Reformulación	30%
		7	Interpretación	70%
22. Emitir juicios sobre un concepto o teoría	10	8	Interpretación	80%
		2	Propósito	20%
23. Indicar acuerdo con un autor o su planteamiento	10	10	Interpretación	100%
24. Sintetizar un planteamiento teórico	10	10	Resumen	100%

Función	Nº de ejemplos	Nº de relaciones RST	Tipo de relaciones RST	%
25. Complementar un planteamiento teórico	10	1	Causa	10%
		1	Concesión	10%
		4	Elaboración	40%
		1	Interpretación	10%
		3	Núcleo (la Causa es el fragmento de la columna uno)	30%
26. Comparar planteamientos teóricos	10	6	Elaboración	60%
		2	Interpretación	20%
		1	Reformulación	10%
		1	Interpretación	10%
27. Aplicar teorías o investigaciones a otros contextos	10	9	Interpretación	90%
		1	Medio	10%
28. Proponer aspectos teóricos	5	1	Propósito	20%
		1	Interpretación	20%
		3	Fondo	60%
29. Destacar la importancia de un concepto o teoría	10	2	Interpretación (+ Causa)	20%
		4	Interpretación	40%
		2	Interpretación (+ Concesión)	20%
		1	Interpretación (+ Antítesis)	10%
		1	Interpretación (+ Reformulación)	10%
30. Indicar opción teórica	10	7	Medio	70%
		2	Secuencia	20%
		1	Fondo	10%
31. Introducir conceptos	8	8	Fondo	100%
32. Exponer hallazgos de la investigación	10	6	Resultados	60%
		1	Resumen (de los Resultados)	10%
		3	Resultados	30%

Función	Nº de ejemplos	Nº de relaciones RST	Tipo de relaciones RST	%
33. Ejemplificar resultados y hallazgos de la investigación	10	9	Elaboración	90%
		1	Evidencia	10%
34. Describir material didáctico	10	10	Medio	100%
35. Realizar interpretaciones o deducciones	10	9	Interpretación	90%
		1	Reformulación	10%
36. Contradecir investigaciones previas	8	5	Antítesis	62,5%
		1	Contraste	12,5%
		1	Interpretación	12,5%
		1	Concesión	12,5%
37. Corroborar investigaciones previas	10	2	Elaboración	20%
		1	Propósito	10%
		1	Lista	10%
		1	Resultados	10%
		4	Interpretación	40%
		1	Antítesis	10%
38. Corroborar hipótesis de investigación	9	4	Evidencia	44,44%
		4	Interpretación	44,44%
		1	Antítesis	11,11%
39. Refutar hipótesis de investigación	10	10	Interpretación	100%
40. Indicar aportes	10	4	Lista	40%
		1	Resultado	10%
		1	Antítesis	10%
		1	Resumen	10%
		1	Interpretación	10%
		1	Elaboración	10%
		1	Secuencia	10%
TOTAL	386	386		