



Universidad de Chile  
Facultad de Filosofía y Humanidades  
Escuela de Postgrado

PLAGIO E INTEGRACIÓN DE FUENTES MÚLTIPLES EN TEXTOS DE  
ESTUDIANTES DE PRIMER AÑO DE UNIVERSIDAD. UNA  
APROXIMACIÓN DESDE LA TEORÍA DE LA COMPRENSIÓN Y EL  
ANÁLISIS SEMÁNTICO LATENTE.

Tesis para optar al grado de Magíster en Lingüística con mención en Lengua  
Española

Alumno: Patricio Moya Muñoz

Director: Guillermo Soto Vergara

Director externo: Juan D. Velásquez Silva

SANTIAGO DE CHILE

2011

## AGRADECIMIENTOS

Al proyecto FONDEF (DO81-1015): DOCODE.

Al profesor Juan D. Velásquez del Departamento de Ingeniería Industrial de la Universidad de Chile por su incondicional apoyo y confianza en mis capacidades.

Al profesor Guillermo Soto del Departamento de Lingüística de la Universidad de Chile por sus sabios comentarios e incesante fomento a la interdisciplina.

A quienes colaboraron en la extracción y presentación de los datos, en especial a Felipe Bravo-Márquez y Gastón L'Huillier Chaparro, ya que sin su desinteresada cooperación esta tarea hubiese sido imposible; y a mi querida amiga Belén Villena Araya por su constante presencia y ayuda en la presentación de este trabajo.

A todo el equipo de desarrollo de DOCODE, en especial a Sebastián Ríos, Héctor Álvarez, Gabriel Oberreuter, Eduardo Merlo, Gerardo Guerrero y Francisco Molina, quienes, a pesar de las diferencias disciplinarias, acogieron, respetaron y apoyaron mi labor dentro del grupo.

A Soledad Chávez por creer en mis capacidades y darme la oportunidad de participar en el grupo DOCODE.

## RESUMEN

El objetivo de este trabajo es investigar sobre la relación que existe entre el *plagio académico* y el nivel de comprensión derivada de la lectura de dos documentos en alumnos de dos carreras diferentes de primer año universitario. La investigación considera la teoría de comprensión de textos planteada, principalmente, por Kintsch (1998), el modelo de documentos postulado por Perfetti, Britt y Rouet (1999) y Rouet 2006, la comparación automatizada de documentos a través de Análisis Semántico Latente (LSA), las formas de transmisión del conocimiento ajeno en el discurso académico, según Hyland (2000) y Bazerman (2004), y el plagio académico presentado por Pecorari (2008).

El corpus empleado corresponde a 43 documentos de alumnos de Licenciatura en Lengua y Literatura y 61 de Plan Común de Ingeniería, ambas de la Universidad de Chile. Estos textos se obtuvieron a partir de una evaluación escrita que formó parte de las actividades académicas de una asignatura en cada disciplina.

Los resultados indican que los textos considerados como *plagios* se presentan, principalmente, en los niveles de comprensión más bajos de cada disciplina. Por el contrario, los documentos que efectivamente atribuyen información a través de citas se ubican entre aquellos que manifiestan una representación de la comprensión de textos más adecuada. De acuerdo con nuestra perspectiva, los estudiantes de primer año universitario no manejan las herramientas necesarias para referirse adecuadamente a la información ajena en la escritura académica formal.

## ÍNDICE GENERAL

AGRADECIMIENTOS	2
RESUMEN	3
ÍNDICE GENERAL	4
ÍNDICE DE CUADROS	7
ÍNDICE DE FIGURAS	9
0. INTRODUCCIÓN	11
1. PLANTEAMIENTO DEL PROBLEMA	13
1.1.OBJETIVO GENERAL	18
1.2. OBJETIVOS ESPECÍFICOS	18
2. MARCO TEÓRICO	19
2.1. REPRESENTACIÓN MENTAL DEL DISCURSO	19
2.1.1. REPRESENTACIÓN MENTAL DE UN DOCUMENTO	21
2.1.1.1. NIVELES DE REPRESENTACIÓN	21
2.1.1.2. ELABORACIÓN DE LAS REPRESENTACIONES DEL DISCURSO A PARTIR DE UN ÚNICO DOCUMENTO	23
2.1.1.3. MODELO DE SITUACIÓN	29
2.1.2. REPRESENTACIÓN DE LA COMPRESIÓN A PARTIR DE FUENTES MÚLTIPLES: MODELO DE DOCUMENTOS	32
2.2. ANÁLISIS SEMÁNTICO LATENTE	34

2.2.1.	EL ANÁLISIS SEMÁNTICO LATENTE Y SU RELACIÓN CON EL MODELO DE KINTSCH	34
2.2.2.	IMPLEMENTACIÓN DEL LSA	36
2.3.	ESCRITURA ACADÉMICA	41
2.4.	DISCURSO AJENO	45
2.4.1.	TRADICIÓN GRAMATICAL	45
2.4.2.	EL DISCURSO AJENO EN LA ESCRITURA ACADÉMICA	48
2.4.3.	PLAGIO ACADÉMICO	51
3.	METODOLOGÍA	59
3.1.	TIPO DE ESTUDIO	59
3.2.	RECOPIACIÓN DE LA INFORMACIÓN	59
3.3.	SUJETOS DE ESTUDIO	60
3.4.	TRATAMIENTO DE LA INFORMACIÓN	60
3.5.	ANÁLISIS DE DATOS	62
4.	ANÁLISIS Y PRESENTACIÓN DE LOS RESULTADOS	64
4.1.	APROXIMACIÓN A LA REPRESENTACIÓN DE LA COMPRESIÓN DERIVADA DE LA LECTURA DE DOS DOCUMENTOS A PARTIR DE LSA	64
4.1.1.	VISUALIZACIÓN DE LA COMPRESIÓN DE DOCUMENTOS	64
4.1.2.	INTEGRACIÓN DE LA INFORMACIÓN Y DISTANCIA TEMÁTICA ENTRE DISCIPLINAS	67
4.2.	INTEGRACIÓN DE LA INFORMACIÓN AJENA: CITAS	71
4.2.1.	ANÁLISIS SINTÁCTICO	72

4.2.2.	ANÁLISIS DE CITAS A PARTIR DEL DISCURSO ACADÉMICO	75
4.3.	OTRAS CARACTERÍSTICAS	81
4.3.1.	INTRODUCTOR DE CITA INTEGRAL	81
4.3.2.	TRATAMIENTO DE LOS REFERENTES	87
4.3.3.	PRESENCIA EXPLÍCITA DE UN APARTADO DE BIBLIOGRAFÍA	90
4.4.	PRESENCIA DE PLAGIO ACADÉMICO	91
4.5.	RELACIÓN ENTRE PLAGIO ACADÉMICO Y REPRESENTACIÓN DE LA COMPRESIÓN DEL DISCURSO	94
5.	CONCLUSIONES	99
6.	REFERENCIAS BIBLIOGRÁFICAS	107
7.	ANEXOS	116

## ÍNDICE DE CUADROS

Cuadro 1: <i>Definición de plagiar en la lengua común: los diccionarios</i>	14
Cuadro 2: <i>Sanciones establecidas para los tipos de plagio</i>	16
Cuadro 3: <i>Ejemplo de representación proposicional de un texto</i>	25
Cuadro 4: <i>Matriz de conectividad para el gráfico de la Figura 2</i>	28
Cuadro 5: <i>Ejemplo de títulos de tópicos sobre música y cocina</i>	38
Cuadro 6: <i>Matriz tipo por documento en relación con los datos del Cuadro 5</i>	38
Cuadro 7: <i>Pesos de la matriz tipo por documento en relación con los datos del Cuadro 6</i>	39
Cuadro 8: <i>Gradación en la presentación del discurso reproducido</i>	46
Cuadro 9: <i>Condiciones caracterizadoras del discurso reproducido</i>	47
Cuadro 10: <i>Técnicas de representación intertextual</i>	49
Cuadro 11: <i>Porcentaje de tipos de citas empleados en las distintas disciplinas</i>	51
Cuadro 12: <i>Categorización de los métodos de identificación de plagio</i>	53
Cuadro 13: <i>Componentes de un archivo XML</i>	61
Cuadro 14: <i>Comparación de las distancias LSA entre los textos de estímulo de cada asignatura</i>	67
Cuadro 15: <i>Cantidad y porcentaje de palabras citadas según token frequency</i>	71
Cuadro 16: <i>Porcentaje de uso de citas directas e indirectas por disciplina</i>	72
Cuadro 17: <i>Ejemplos de cita directa e indirecta en las disciplinas</i>	73
Cuadro 18: <i>Porcentaje de tipos de citas directas con y sin atribución</i>	74
Cuadro 19: <i>Ejemplos de cita directa sin estrategia sintáctico/textual introductora</i>	74
Cuadro 20: <i>Caracterización de la gradación de las referencias propias de las citas</i>	76
Cuadro 21: <i>Ejemplos de tipos de citas integrales y no integrales por disciplina</i>	79

Cuadro 22: <i>Verbos introductores de citas integrales en LLyLH</i>	83
Cuadro 23: <i>Verbos introductores de citas integrales en PCI</i>	84
Cuadro 24: <i>Otros recursos para la integración de la voz ajena en LLyLH</i>	85
Cuadro 25: <i>Otros recursos para la integración de la voz ajena en PCI</i>	85
Cuadro 26: <i>Ejemplos de recursos empleados para integrar la voz ajena</i>	86
Cuadro 27: <i>Recursos para el tratamiento de referentes en LLyLH</i>	87
Cuadro 28: <i>Recursos para el tratamiento de referentes en PCI</i>	88
Cuadro 29: <i>Ejemplos de recursos para el tratamiento de referentes</i>	89
Cuadro 30: <i>Promedios de uso LSA en textos plagiados en LLyLH</i>	94
Cuadro 31: <i>Promedios de uso LSA en textos plagiados en PCI</i>	95
Cuadro 32: <i>Porcentaje de distribución de textos plagiados según distancia LSA en LLyLH</i>	96
Cuadro 33: <i>Porcentaje de distribución de plagio según distancia LSA en PCI</i>	97
Cuadro 34: <i>Porcentaje de distribución de citas según distancia LSA en LLyLH</i>	97
Cuadro 35: <i>Porcentaje de distribución de citas según distancia LSA en PCI</i>	98



## ÍNDICE DE FIGURAS

Figura 1: <i>Esquema de diferenciación entre texto base y modelo de situación</i>	22
Figura 2: <i>Conexiones entre BANCO 1 y BANCO 2 y sus asociados</i>	28
Figura 3: <i>Representación de la matriz ortogonal</i>	39
Figura 4: <i>Representación de SVD</i>	41
Figura 5: <i>Modelo de Flowers y Hayes (1981)</i>	43
Figura 6: <i>Gráfico de representación del objetivo de la tarea en relación con el plagio</i>	54
Figura 7: <i>Esquema de las cuatro formas en que los estudiantes cometen plagios</i>	56
Figura 8: <i>Distancia entre los documentos en LLyLH</i>	65
Figura 9: <i>Distancia entre los documentos en PCI</i>	66
Figura 10: <i>Gráfico de porcentajes de distribución de documentos según distancias coseno LSA, LLyLH</i>	68
Figura 11: <i>Gráfico de porcentajes de distribución de documentos según distancias coseno LSA, en PCI</i>	69
Figura 12: <i>Gráfico de porcentajes de distribución de documentos según promedio de distancia coseno LSA entre los documentos</i>	70
Figura 13: <i>Gráfico de porcentaje de tipos de citas según la clasificación planteada por Swales (1990) ampliada según el tipo de referencia implicada</i>	77
Figura 14: <i>Gráfico de porcentaje de tipo de introductor de cita integral</i>	82
Figura 15: <i>Gráfico de porcentaje de documentos con bibliografía explícita en ambas disciplinas</i>	90
Figura 16: <i>Gráfico de porcentaje de textos plagiados en LLyLH y PCI siguiendo a Pecorari (2008)</i>	91

Figura 17: *Gráfico de presencia de plagio, de acuerdo con criterios definitorios para alumnos de primer año* 93

Figura 18: *Gráfico de comparación de promedio de distancias LSA en textos plagiados en LLyLH y PCI* 96

## 0. INTRODUCCIÓN

El estudio de la comprensión del discurso ha descansado, principalmente, en la caracterización de la representación cognitiva que realizan las personas cuando leen un documento. En este ámbito, los trabajos de Kintsch (1988, 1996, 1998) y Kintsch y Van Dijk (1978, 1983), centrados en la descripción de los procesos de construcción e integración de la información, han tenido gran impacto dentro de la comunidad científica. La comprensión de un texto se identifica con la elaboración de un *modelo situacional* de los hechos que en él se presentan y se almacena en la memoria de largo plazo.

Sin embargo, la mayor cantidad de evaluaciones y tareas de escritura en la enseñanza superior requieren de la lectura de más de un documento, por lo que el modelo presentado por Kintsch no se adecuaría a estos requerimientos. A partir de lo anterior, Perfetti, Britt y Rouet (1999) plantean la necesidad de considerar un modelo que se corresponda con los factores que inciden en la comprensión de múltiples documentos (denominado como *modelo de documentos*), como, por ejemplo, la relación entre los modelos situacionales derivados de cada texto.

Por otra parte, las investigaciones relativas a la escritura académica se han ocupado, casi exclusivamente, de la especificación de los diversos rasgos que articulan la transmisión del conocimiento disciplinario en artículos de carácter científico, escritos por especialistas en las materias. Además, estas caracterizaciones se han realizado, en mayor medida, para el idioma inglés (Hyland, 2002). Dentro de estas características, las estrategias para transmitir el conocimiento ajeno (Bazerman, 2004) se constituyen como uno de los aspectos más importantes para que el discurso se valide dentro de la comunidad discursiva en la que se adscribe. Dichas estrategias implican la atribución de la información, con el fin de identificar su autoría. En este sentido, el estudio sobre el *plagio* (Pecorari, 2008; Sutherland-Smith, 2008), entendidos como secuencias textuales provenientes de otra fuente sin una adecuada referencia acerca de su procedencia, se ha enfocado en determinar los rasgos relacionados tanto con la escritura académica, como con las diferencias culturales en torno al concepto de autoría.

La presente investigación se propone investigar la relación entre comprensión de textos derivada de la lectura de dos documentos y los mecanismos de transmisión del conocimiento ajeno, en estudiantes de primer año universitario. Estamos conscientes de la importancia que ha tomado la enseñanza de habilidades escriturales a los estudiantes, en las diversas instituciones de educación superior. Por lo anterior, los resultados de este trabajo pueden aportar en la elaboración de los diversos programas de las asignaturas de expresión escrita, al caracterizar las habilidades de referencia al conocimiento ajeno por parte de alumnos que se encuentran en la etapa inicial de acceso a la comunidad discursiva.

## 1. PLANTEAMIENTO DEL PROBLEMA

La derivación etimológica de la voz *plagio* (del latín *plagiare*) se considera como un cultismo, debido al ámbito en el que se usaba, antiguamente, con frecuencia: el judicial. En la Edad Media, guardaba, principalmente, el significado de “el raptó del hombre, la compra o venta de una persona libre como esclava<sup>1</sup>” (Colón, 1992, p. 9). Si bien *Plagio* con el sentido de ‘hurto intelectual’ se utilizaba ya en la Antigua Roma, no es sino hasta el siglo XVIII cuando la mayor parte de los idiomas romances dieron este sentido a la voz. En español, el *Diccionario castellano con las voces de ciencias y artes* de Esteban Terreros y Pandó (1786-1793) fue el primer repertorio lexicográfico en lengua española que integró la voz *plagiar*, de la cual señala que significa “hurtar los pensamientos ajenos para publicarlos por propios, usar de las obras de otro acomodándolas a sí mismo. (...). Aunque no hay modo alguno noble de hurtar, el plagiar es muy villano”.

Con el sentido de apropiación de una idea ajena, en la actualidad, el *plagio* (o *plagiar*) se ha transformado en un grave problema para la sociedad en general, originado, en gran parte, en las facilidades tecnológicas que existen para acceder y operar con diferentes fuentes de información. Uno de los acercamientos hacia la definición del concepto de *plagio* se relaciona con el conocimiento depositado en los diferentes diccionarios de lengua, esto es, el tipo de información a la que las personas acuden como primera fuente de aproximación.

Cuadro 1: *Definición de plagiar en la lengua común: los diccionarios*

Diccionario	Definición
<b>DRAE</b>	Copiar en lo sustancial obras ajenas, dándolas como propia.
<b>Merriam-Webster</b>	To steal and pass off (the ideas or words of another) as one's own: use (another's production) without crediting the source.
<b>Oxford</b>	1. Take (the work or an idea of someone else) and pass it off as one's own. 2. Copy from (someone) and pass it off as one's own.

*Nota 1.* Adaptado de DRAE (2001), Merriam – Webster (2010) y Oxford University Press (2010).

Las definiciones presentadas en el Cuadro 1 se distancian entre sí, evidenciando el poco consenso que existe en su conceptualización. En primer lugar, el DRAE señala que *plagiar* se relaciona con la reproducción de obras ajenas y no especifica la cantidad de información que constituye una *obra*. Al respecto, las definiciones presentadas tanto en Merriam-Webster como en Oxford concuerdan en que lo *plagiado* pueden ser tanto palabras como ideas. En segundo lugar, difieren en la *acción* que es considerada como *plagiar*: mientras para DRAE y el Oxford Dictionary constituye una *copia*, para el Merriam-Webster Dictionary se establece como un *robo*. Por último, solamente el Merriam-Webster Dictionary se detiene en que *plagiar* no tan solo implica hacer pasar cierto segmento textual como propio, sino que especifica la inexistencia de una referencia que indique su procedencia.

Académicamente, el plagio es considerado como parte de una serie de *engaños* o *deshonestidades académicas* tales como: copiar en pruebas y tareas, falsificar los datos en investigaciones, usar inapropiadamente recursos o manipular equipamiento académico no pertinente (Park, 2003, p. 474). En los reglamentos estudiantiles universitarios, se verifica un segundo acercamiento al concepto del *plagio*. En el caso de las dos universidades chilenas más prestigiosas<sup>2</sup> (Universidad de Chile y Pontificia Universidad Católica de Chile), no existe consenso ni en la conceptualización ni en las penas académicas que involucran<sup>3</sup>. El reglamento interno de pre-grado de la Universidad de Chile (1993) no

menciona la palabra *plagio* y señala que los estudiantes deben “reconocer el origen y autoría de las ideas y resultados tanto propios como ajenos, según las normas y convenciones académicas de cada disciplina”; además, no puntualiza sanciones relacionadas con el incumplimiento de este deber estudiantil. Por su parte, el reglamento de los estudiantes de pre-grado de la Pontificia Universidad Católica (2008) especifica que “plagiar u ocultar intencionalmente el origen de la información en investigaciones y trabajos en general” se debe considerar como una falta a la honestidad académica, cuya sanción implica la nota mínima. La preocupación actual radica en la necesidad de especificar lo que se considera un plagio o falta a la honestidad académica, puesto que el principal problema que tienen las instituciones en la actualidad es tratar con una *generación sin miedo*, que creen que copiar material proveniente de la *Web* es adecuado y pertinente para las necesidades académicas (Park, 2003, p. 475).

El fenómeno del *plagio* se relaciona, también, con las leyes de propiedad intelectual (en Chile, la ley 17.336). A diferencia de las definiciones entregadas por los diccionarios, la ley 17.336 se preocupa no solo de los textos escritos, sino que, además, se encarga de resguardar los intereses de los autores de diferentes tipos de obras, independientes del carácter físico que en el que se manifiesten (pintura, programas computacionales, grabaciones musicales, obras cinematográficas, etc.). En Chile, la ley –promulgada en 1970 y modificada por última vez el año 2004–, tiene por objetivo proteger

los derechos que, por el solo hecho de la creación de la obra, adquieren los autores de obras de la inteligencia en los dominios literarios, artísticos y científicos, cualquiera que sea su forma de expresión, y los derechos conexos que ella determina.

A pesar de no mencionar en ninguno de sus incisos la palabra *plagio*, identifica, explícitamente, una serie de factores que determina el empleo inadecuado de las fuentes originales, entre las que se puede mencionar: el uso exclusivo por parte del titular de la obra; la autorización directa para el uso de la obra depende del titular; los derechos se mantienen 70 años después del fallecimiento del titular; entre otros. Según la ley, se considera como falta o delito a quien “sin estar expresamente facultado para ello, utilice

obras de dominio ajeno protegidas por esta ley, inéditas o publicadas”. Las sanciones que corresponden para este delito dependen del monto del perjuicio, esto es, mientras mayor es el monto del perjuicio mayor será la sanción (con pena de cárcel o multa).

En general, las sanciones establecidas para el *plagio* dependerán del contexto en el que se produzca. Dichos contextos se identifican con los tres acercamientos revisados (ver Cuadro 2).

Cuadro 2: *Sanciones establecidas para los tipos de plagio*

<b>Sanción</b>	<b>Tipo de plagio</b>	<b>Caracterización</b>
Informal, no legal: estigma social	Relacionado con los medios de comunicación escritos.	Los que violan las reglas de atribución propias de la comunidad son desestimados por sus pares.
Formal, cuasi legal: con procedimientos académicos y disciplinarios	Relacionado con instituciones educativas.	Existe una “justicia privada” que se ocupa de solucionar los casos de plagio. La comunidad, implícita o explícitamente, acepta las reglas impuestas.
Legal	Relacionado con la propiedad intelectual.	Implica multas o penas de cárcel, por sacar provecho económico de sus actos.

*Nota 1.* Basado en Green (2002).



La presente investigación se inscribe dentro de los estudios sobre el *plagio* producido en las instituciones educativas, en otras palabras, el denominado *plagio académico*. Los trabajos en el área han puesto su atención en tres aspectos fundamentales. Primero, los estudios definicionales se aproximan al hecho mediante la revisión de las nociones que tienen los estudiantes sobre el concepto del *plagio* (por extensión, se preocupan de la intencionalidad del hecho); luego, los estudios constructivo-culturales, que examinan los factores culturales determinantes en las características que manifiestan los alumnos al momento de redactar un trabajo; y, finalmente, los estudios textuales que operan sobre las prácticas de escritura, incluyendo tanto los aspectos técnicos del sistema de referencias bibliográficas, como las estrategias generales que involucran la transmisión del conocimiento ajeno (Sutherland-Smith, 2008). Muchas veces esta diferenciación no pasa de ser metodológica, en la medida en que numerosas pesquisas correlacionan los datos derivados de más de un enfoque, por ejemplo, entre las formas de citación y el país de origen de los estudiantes (cf. Pecorari, 2003).

Sin embargo, son escasos los estudios que relacionan el fenómeno del plagio académico con la comprensión de lectura alcanzada por los estudiantes. Al respecto, Bhattacharya y Jorgensen (2008) precisan que la mejor manera para reducir los casos de plagio al interior de las aulas se identifica con el control que deben realizar los profesores sobre la información que utilizan los estudiantes. De esta manera, las tareas de escritura enviadas a los estudiantes deben promover el aprendizaje de los contenidos, esto es, la comprensión efectiva de los textos.

## 1.1. OBJETIVO GENERAL

Determinar la relación que existe entre las estrategias de transmisión del conocimiento ajeno, incluido el plagio, y la representación de la comprensión de múltiples documentos en estudiantes de primer año universitario en dos carreras de la Universidad de Chile.

## 1.2. OBJETIVOS ESPECÍFICOS

- a. Describir, a través de análisis semántico latente, el nivel de representación de la comprensión de los textos de los estudiantes de primer año universitario a partir de la lectura de dos fuentes.
- b. Caracterizar las estrategias de transmisión del conocimiento ajeno, incluido el plagio, utilizadas por los estudiantes de primer año universitario.

## 2. MARCO TEÓRICO

### 2.1. REPRESENTACIÓN MENTAL DEL DISCURSO

La comprensión de los discursos (Belinchón, Igoa, Rivière, 2000, pp. 469-471) no se centra exclusivamente en la decodificación de cada una de las oraciones que los componen, puesto que todas las personas buscan dotar a los textos de coherencia, tarea imposible de realizar si se ciñeran, exclusivamente, a las proposiciones explícitas que los forman. Los textos no se recargan de datos que puedan ser deducibles (es decir, inferibles) de acuerdo al conocimiento de mundo que posean los participantes. De esta manera, tal como señalan Belinchón, Igoa y Rivière (2000, p. 477):

El plano discursivo constituye el punto de encuentro principal entre los mecanismos más específicamente lingüísticos (tales como los que permiten procesar la sintaxis oracional) y los sistemas de pensamiento más globales, que se basan en procedimientos generales de inferencia y en bases también generales de conocimientos sobre el mundo, que están representados en el sistema cognitivo.

Para Graesser, Gernsbacher y Goldman (2000, pp. 427-428), los modelos cognitivos del discurso han estado influenciados tanto por las teorías cognitivas simbólicas como por las conexionistas. En las primeras, existe, por un lado, una memoria de trabajo y, por otro, una reserva de proposiciones, esquemas y miles de reglas de producción, del tipo SI[condiciones] ENTONCES[acción (física o cognitiva)], que se evalúan constantemente en la medida en que ingresa información nueva a la memoria de trabajo. En las teorías conexionistas, en cambio, tanto procesos como representaciones se encuentran organizados en un gran conjunto de unidades simples, denominadas *unidades neuronales*. Así, las palabras, conceptos, proposiciones o esquemas poseen un conjunto de unidades que se conectarán entre sí por pesos y forman, en consecuencia, redes neuronales. Tal como señala Carreiras (1997), a diferencia de los modelos simbólicos, los conexionistas “son enseñados más que programados, ya que utilizan mecanismos de aprendizaje poderosos que, por exposición a muchos ejemplos, logran crear un grupo de representaciones internas

complejas” (p.159). Se establecen como sistemas dinámicos<sup>4</sup> que permiten la aparición de representaciones graduales, lo que ayuda en el modelamiento del proceso de aprendizaje (p. 160).

La mayor parte de los estudios sobre comprensión del discurso ha descansado en la investigación sobre la representación mental de un único documento. Dentro de esta perspectiva, los trabajos realizados por Kintsch y Van Dijk (1978, 1983) y Kintsch (1988, 1994, 1996, 1998) – adoptada en esta investigación – siguen vigentes y han trascendido a una gran cantidad de investigaciones posteriores. Además, para los fines perseguidos en el presente trabajo, resulta fundamental considerar la comprensión a partir de la lectura de más de un documento (Perfetti, Rouet y Britt, 1999; Cerdán, 2006; Rouet, 2006; Vega, 2011), puesto que, si bien comparte un sustento teórico común sobre el tratamiento de los textos de manera individual, se distancia de él en la medida en que se relaciona de mejor manera con las diversas tareas impuestas desde las instituciones educativas.

## 2.1.1. REPRESENTACIÓN MENTAL DE UN DOCUMENTO

### 2.1.1.1. NIVELES DE REPRESENTACIÓN

Se reconocen tres niveles la representación de un texto (Kintsch y Van Dijk, 1983; Kintsch, 1994; Kintsch, 1996, Kintsch, 1998; Kintsch y Kintsch, 2005; Vega, Díaz y León, 1999). El proceso de decodificación de la *estructura de superficie* implica los procesos conceptuales y perceptuales involucrados en la obtención de los significados desde la letra impresa hacia la mente del lector. Un segundo nivel de análisis se compone tanto de de la *microestructura* del texto (una compleja red conectada de proposiciones, en otras palabras, la información en y entre las oraciones) como de la *macroestructura* (relaciones globales jerárquicas entre varios segmentos del texto) (Kintsch y Kintsch, 2005, p.73). En conjunto, ambas constituyen el *texto base* (Kintsch, 1998, pp.103-105), que representa el significado de un texto tanto a nivel local como global. Se identifica con la representación mental de la parte que se deriva directamente de los textos, cuya construcción se realiza sobre la base de una gran cantidad de conocimiento, tanto lingüístico como del mundo; no obstante, este conocimiento se ocupa, exclusivamente, para comprender los contenidos implicados en el texto (Kintsch, 2002).

Finalmente, como se detallará en 2.1.1.3., la información proveniente del texto base debe integrarse junto con el conocimiento previo del lector, con el fin de construir un *modelo de situación*, esto es, la representación cognitiva de los eventos, acciones y la situación en general del texto. Con todo, esta distinción –en especial la establecida entre texto base y modelo de situación ejemplificada en la Figura 1– es netamente metodológica, de acuerdo a lo señalado por Kintsch (1998, p. 107):

We distinguish between textbases and situation models not because are somehow different mental objects. On the contrary, there is a single, unitary mental representation of a text. To describe this representation and its genesis, it is useful to separate out

textbases and situation models, as well as problem models and surface structures, as our analysis require.

Figura 1: Esquema de diferenciación entre texto base y modelo de situación<sup>5</sup>

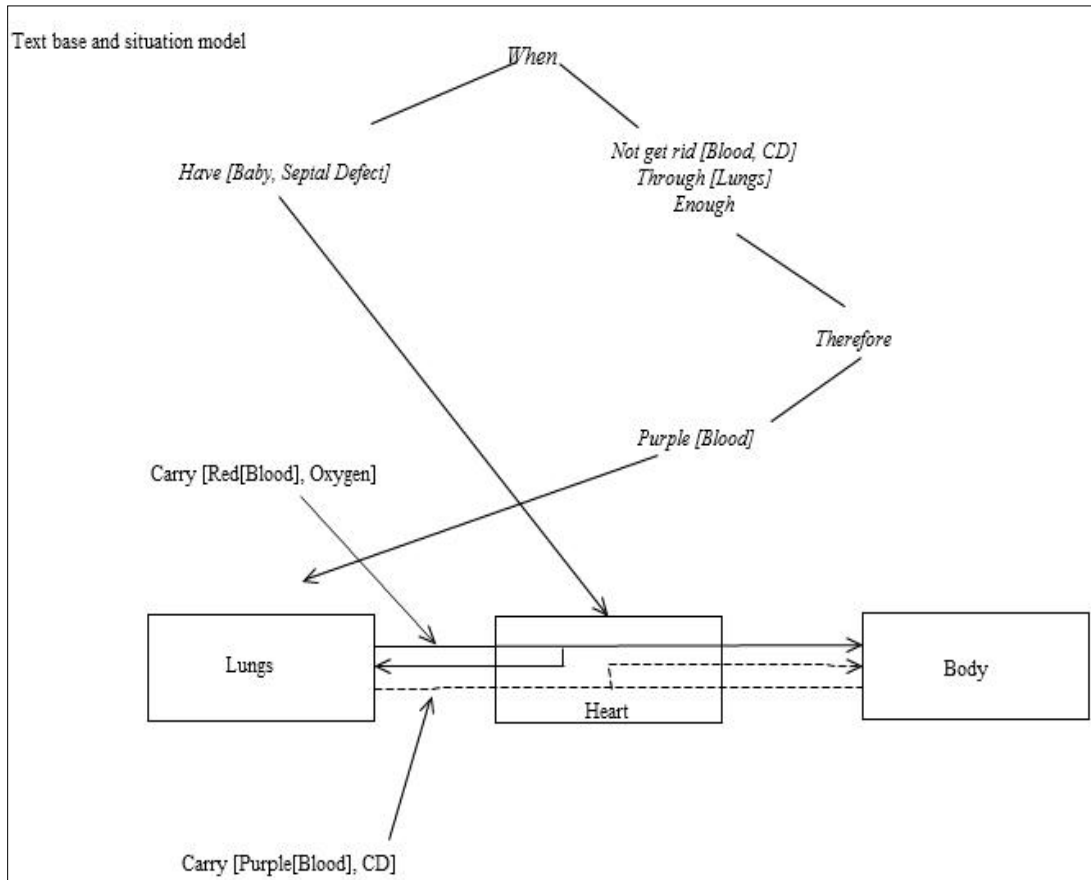


Figura 1. Esquema planteado a partir del texto: “When a baby has a septal defect, the blood cannot get rid of enough carbon dioxide through the lungs. Therefore, it looks purple”. El texto base se señala por medio de letra cursiva y se conecta con nodos de conocimiento previo sobre el tema. Esta relación permite establecer el modelo situacional del texto.

La comprensión de textos se encuentra íntimamente relacionada con la representación sobre un texto que logran las personas. Para Kintsch, comprender un discurso, desde la teoría de la computación, significa “construir la representación de un discurso sobre el cual pueden realizarse varios cálculos, cuyos resultados se toman como evidencia de la comprensión” (1996, p.71). De lo anterior, se desprende que existirán representaciones más adecuadas que otras, tal como afirma Rouet (2006, p.6) “comprehending a text is a process in which the reader attempts to mentally represent the situation described in the text”.

#### 2.1.1.2. ELABORACIÓN DE LAS REPRESENTACIONES DEL DISCURSO A PARTIR DE UN ÚNICO DOCUMENTO

La primera versión de la *teoría macroestructural* (cf. Vega, Díaz y León, 1999, p. 280) presentada por Kintsch y Van Dijk (1978, 1983) se destaca por el carácter dinámico de la comprensión lectora, en tanto se considera que el procesamiento del discurso se realiza en ciclos, determinados por la capacidad de almacenamiento de la memoria de corto plazo. Además, se concibe como un modelo *estratégico*, en la medida en que se aplican diversas *estrategias*<sup>6</sup> al proceso de lectura, dependientes no solo de las características textuales, sino que también de las metas o el conocimiento de mundo de los lectores. La estrategia general, desde la cual se derivan las demás, es la que busca la construcción del *texto base*, que se cumple cuando se satisface una cantidad mínima de criterios relacionados con otorgar coherencia local y global al texto (Kintsch y Van Dijk, 1983, pp. 10-11).

Dichas coherencias se establecen a partir de la noción de que la estructura de superficie de un discurso se debe interpretar como un conjunto de proposiciones, ordenadas según una serie de relaciones semánticas derivadas de su propia interconexión (Kintsch y Van Dijk, 1978). Esta estructura semántica está compuesta por dos niveles: la *microestructura* y la *macroestructura*. La microestructura es la estructura compuesta por las proposiciones individuales y las relaciones resultantes entre ellas (Kintsch, 1978, p. 364), en otras palabras, se construye a partir del texto superficial con el objetivo de

establecer la correferencia. Por su parte, la macroestructura considera el texto como un todo, con el fin de retener su sentido (Vega, Díaz y León, 1999, p. 281). Para lograrlo, se construye a partir de los datos proporcionados por la microestructura, cuya información es reducida por medio de *macrorreglas* que aseguran la extracción del tópico o tema del discurso. Las macrorreglas descritas (Kintsch y Van Dijk, 1978; Van Dijk, 1992) son:

- a. Omitir: se deben eliminar todas las proposiciones que no se relacionen directa o indirectamente con las condiciones de la interpretación impuestas por el lector
- b. Generalizar: se debe sustituir una secuencia de proposiciones por una sola proposición que contenga el contenido semántico del conjunto reemplazado.
- c. Construir: se debe sustituir una secuencia de proposiciones que denoten hechos por una sola proposición que denote ese hecho global.

Cabe destacar que estas macrorreglas no son solo operaciones mentales aisladas, vale decir, que se producen gracias a un conocimiento general, sino que dependen de los conocimientos previos que tenga cada lector. Estos conocimientos previos se representan por medio de estructuras discursivas esquemáticas<sup>7</sup> que cumplen un rol preponderante en la comprensión discursiva, dado que sin ellas las personas no serían capaces, por ejemplo, de juzgar si un argumento es correcto en un determinado discurso (Kintsch y Van Dijk, 1978, p. 366). Asimismo, ayudarían a determinar y guiar las potenciales posibilidades la lectura:

Un hablante solo será capaz de decidir qué tipo de información seguirá siendo probablemente importante en el texto o qué tipo de circunstancia global se describe en el texto, cuando las proposiciones añadidas se comparen con proposiciones en las situaciones de marco habitualmente esperadas. (...) La comparación de proposiciones provenientes del texto con componentes del marco de conocimientos no sólo da como resultado los conceptos característicos de este marco (...) sino que a la vez produce una serie de *expectativas* sobre el decurso ulterior de los sucesos y sobre el *posible* desarrollo ulterior del texto (Van Dijk, 1992, p. 200).



Dado que la unidad de análisis es la proposición, el *input* del modelo consiste en un listado de proposiciones del texto que representan su significado. Las proposiciones están formadas por *predicados* –compuestos por conceptos– (o *concepto relacionador*) y uno o más *argumentos* (que puede ser otro concepto o una proposición incrustada). Estos últimos cumplen diferentes funciones semánticas, como agente, objeto o meta; mientras que los predicados se realizan en la estructura de superficie como verbos, adjetivos o adverbios. En cuanto a su presentación, las proposiciones deben establecerse de acuerdo al orden de aparición en el mismo texto (ver Cuadro 3).

Cuadro 3: *Ejemplo de representación proposicional de un texto*

Oración	Proposiciones
The first of the heart's four chambers, the right atrium, receives purplish blood, short of oxygen and laden with carbon dioxide.	1. FIRST (CHAMBER)
	2. POSSESS (HEART, 3)
	3. FOUR (CHAMBER)
	4. RIGHT (ATRIUM)
	5. IS (1, 4)
	6. PURPLE (BLOOD)
	7. RECEIVES (1, 6)
	8. SHORT OF (6, OXYGEN)
	9. LADEN (6, CARBON DIOXIDE)

*Nota 1.* Extraído de Rouet (2006, p. 4).

Luego de la aparición de la primera versión de la *teoría macroestructural*, Kintsch (1988, 1996, 1998) presenta su modelo de *Construcción-Integración*, en el que si bien conserva aspectos esenciales del modelo anterior, como la mantención del formato proposicional y la concepción del proceso en dos etapas (Vega, Díaz y León, 1999, p. 281), varía en el papel que cumple la representación del conocimiento general en los procesos implicados en la comprensión de los discursos. Además, señala que un fragmento de discurso puede ser interpretado de dos maneras diferentes: como una red de conocimientos generales y como la base proposicional de un texto breve. Por lo anterior, Kintsch (1996, p.

77) indica que “los elementos a partir de los cuales se construyen las redes de conocimientos y las bases de texto son los mismos”.

Kintsch es consciente de la dificultad operacional de su concepción anterior, al percatarse de la gran cantidad de pasos alternativos que puede tener el proceso de representación discursiva. En este sentido, aunque el conocimiento<sup>8</sup> restringe las representaciones, no puede seguirse considerando como estructuras esquemáticas, puesto que resultan poco flexibles para adecuarse a las variaciones inherentes al proceso lector (1996, p. 74). En esta concepción, por tanto, se considera que el conocimiento se encuentra escasamente organizado y se activará en la medida en que la lectura así lo requiera. Su representación, esta vez, se efectúa a través de una red asociativa –denominada *red de conocimiento*– (Kintsch, 1996, 1998), en la que los nodos, interconectados entre sí, se corresponden con conceptos o proposiciones<sup>9</sup>. Los nodos (proposiciones o esquemas completos) se componen de un núcleo principal más una cantidad indeterminada de casilleros para los argumentos y sus conexiones se especifican a través de valores de fuerza que van de 1 a -1. El significado de los nodos se establece por su posición dentro de la red de conocimiento, dada por la fuerza con la que se relacionan con los nodos vecinos (Kintsch, 1998, p.74).

Para Kintsch (1998, p. 76), las *redes de conocimiento* permiten dar luces sobre la representación mental del significado de las *palabras* (en la medida en que se consideran como conceptos que pueden articular una *red*). El significado se construye a partir de la activación de los nodos vecinos que se encuentran alrededor de una palabra. De lo anterior, se infiere que el significado de las palabras depende del contexto, por lo que se considera que es dinámico y fluctuante, en tanto “the discourse context is in continuous flux, and different persons operate with different knowledge nets. Hence, there must be considerable variability in effective word meanings” (Kintsch, 1998, p. 77).

En este modelo, las representaciones mentales están formadas por débiles reglas de producción y se desarrollan como un desordenado, redundante e incluso contradictorio *output* (Kintsch, 1998, p. 94), proceso denominado como de *construcción*. Posteriormente, este *output* experimenta un proceso de *integración*, que origina una representación mental

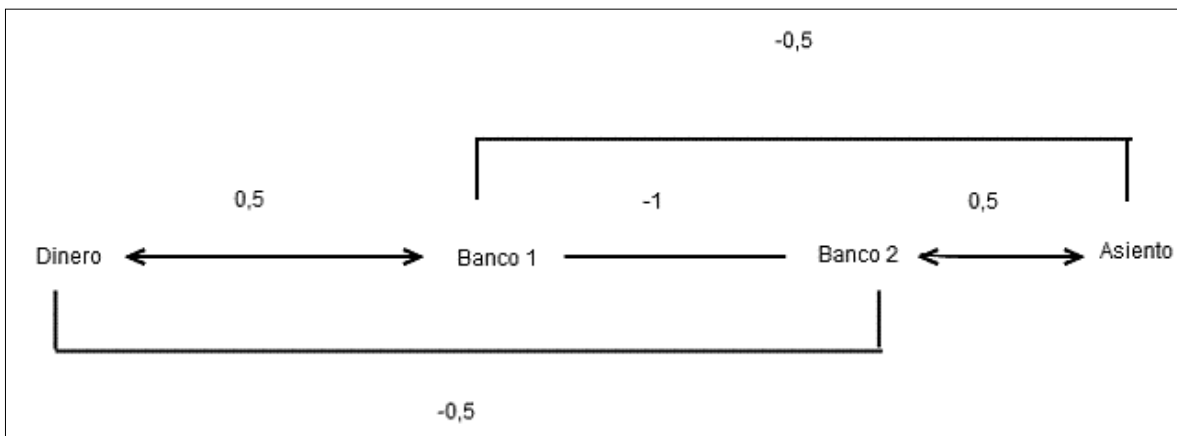
bien estructurada. Específicamente, el proceso de *construcción* implica una serie de reglas o pasos necesarios para su consecución (Kintsch, 1988, p. 81; Kintsch, 1998, p. 96): en primer lugar, se deben formar los conceptos que se refieren al *input* lingüístico (que permite la elaboración de conceptos debilitados, que originan conceptos incompletos o erróneos); luego, cada concepto formado ayuda en la recuperación de la información de los nodos asociados a su propia red de conocimiento. Esta relación entre los nodos puede ser directa, indirecta o subordinada (Kintsch, 1998). En tercer lugar, las reglas para la activación del conocimiento implican que en la memoria de trabajo se activen los nodos vecinos en la red de conocimiento con una probabilidad proporcional a la fuerza con la que cada uno está vinculado con los otros (Kintsch, 1998, p. 97). Por último, se requiere la generación de inferencias específicas y controladas, como las inferencias *punte* en el caso de que el texto base construido sea incoherente (Kintsch, 1996, p. 85). En la Figura 2, se ejemplifica la porción de una red, al activarse la palabra *banco*. Derivada de esta representación, en el Cuadro 4 se presentan los posibles valores de distancia que existirían entre las proposiciones.

Cuadro 4: *Matriz de conectividad para el gráfico de la Figura 2*

Proposición	1	2	3	4
1. Dinero	-	0,5	-0,5	0,0
2. Banco 1	0,5	-	-1,0	-0,5
3. Banco 2	-0,5	-1,0	-	0,5
4. Asiento	0,0	-0,5	0,5	-

Nota 1. Extraído de Kintsch (1996, p. 87).

Figura 2: *Conexiones entre BANCO 1 y BANCO 2 y sus asociados*<sup>10</sup>



Por otra parte, la representación del discurso debe ser concebida como una elaboración secuencial, puesto que es imposible *construir e integrar* un libro completo, por ejemplo: los textos tienen que procesarse en ciclos por palabras y oraciones (Kintsch, 1998, p.101). De esta manera, en tanto un segmento de texto es procesado, se *integra* inmediatamente al resto de la representación del texto. El proceso de *integración* se lleva a cabo cuando la red es construida. Luego, los nodos se activan hasta que se logra la estabilización de la red mediante la multiplicación del *vector de activación* (que se corresponde con los valores iniciales de todos los nodos de activación de la red) con la matriz de conectividad. Con esto, el resultado del proceso de *integración* se constituye en un nuevo vector de activación, que establece valores altos para los nodos que se identifican

con la representación adecuada del discurso y bajos (incluso 0) para aquellos que no forman parte de dicha representación (Kintsch, 1996, p. 90).

El modelo de *Construcción-Integración* ha sido criticado por las limitaciones impuestas por el formato proposicional para dar cuenta de todos los fenómenos involucrados en la comprensión. Se señala que no se especifica la construcción de la *red de conocimiento* derivada de las proposiciones de los textos: no se presenta un detalle sobre la manera en que se produce el paso de la superficie textual a *texto base*, tarea que depende, exclusivamente, del investigador (Vega, Díaz y León, 1999, p. 281). Además, el mismo Kintsch (1998) se ha referido a una de las limitaciones fundamentales del modelo, esto es, su incapacidad para operar con una gran cantidad de textos. Con el objetivo de superar esta dificultad, propone un análisis matemático denominado Análisis Semántico Latente (LSA por sus siglas en inglés) (ver apartado 2.2.2.).

#### 2.1.1.3. MODELO DE SITUACIÓN

La explicitación detallada de los modelos situacionales<sup>11</sup> es necesaria para los objetivos de la presente investigación, puesto que se corresponde directamente con la mantención de la información leída, gracias a la fuerza de los vínculos creados con los conocimientos previos de las personas (por lo que es almacenada en la memoria de largo plazo). De esta manera, es posible recurrir a ellos en contextos diferentes a los que los originaron, facilitando los procesos de aprendizaje de los estudiantes:

The textbase consists of those elements and relations that are directly derived from the text itself (...) this procedure yields an impoverished and often even incoherent network. The reader must add nodes and establish links between nodes from his or her own knowledge and experience to make the structure coherent, to complete it, to interpret it in terms of the reader's prior knowledge, and last but not least to integrate it with prior

knowledge (...) a situation model is, therefore, a construction that integrates the textbase and relevant aspects of the comprehender's knowledge (Kintsch, 1998, pp. 103, 107).

Los modelos situacionales son representaciones similares a las construidas con la experiencia directa con el entorno: no se establecen solamente con las entradas lingüísticas, sino que también intervienen condicionantes espaciales, temporales, causales e, incluso, interpersonales determinadas por el mundo real. Corresponden a representaciones *corporeizadas*, puesto que conservan aspectos propios de la interacción sensorio-motora con las situaciones (Vega, Díaz y León, 1999). De ahí se considera su diferencia fundamental con las estructuras esquemáticas (Kintsch y Van Dijk, 1983, p. 344), vale decir, dependen del conocimiento particular de un sujeto. Su elaboración incluye, por un lado, el texto base y, por otro, largas elaboraciones de conocimiento del mundo y de conocimiento basado en las interpretaciones del discurso. Sin embargo, a pesar de construirse sobre el texto base, no contiene toda su información, ya que selecciona aquellos nodos que le permiten establecer conexiones efectivas con su conocimiento previo sobre el tema, omitiendo a los que se encuentran en posiciones marginales de la red, es decir, reduciendo datos (Zwaan y Radvansky, 1998; Vega, Díaz y León, 1999).

En la *red de conocimiento* elaborada a partir de la lectura del texto, el lector debe añadir nodos y establecer vínculos entre los nodos de su propio conocimiento y experiencia para lograr una estructura coherente. Dado que se relaciona con los conocimientos previos de los sujetos, el tipo de información que opera en la elaboración de los modelos situacionales puede ser sobre el lenguaje, sobre el mundo en general y aspectos específicos de la situación comunicativa (Kintsch, 1998, p.103). En este sentido, en todo texto completo se describe un modelo de situación. En muchas ocasiones, los textos se presentan de forma incompleta y descansan en el supuesto de que el lector completará los vacíos informativos con su conocimiento previo. Pese a lo anterior, si el lector no posee conocimiento previo relevante sobre el contenido o no se ocupa de comprender a cabalidad el texto, la representación definitiva del mismo estará dominada por el texto base. Se

desprende que dos lectores no se formarían el mismo modelo de situación de un mismo texto (Kintsch, 1994).

Junto con lo anterior, Zwaan y Radvansky (1998), basados en Johnson-Laird (1983), indican que el modelo situacional es *multidimensional*, en la medida en que está compuesto por cinco dimensiones: espacial, causal, intencional, sobre los protagonistas y objetos y temporal. Ninguna de ellas tiene un valor superior a otra *per se*, sino que su peso dependerá de la tarea a la que la persona esté sometida (p. 178). En este sentido, por ejemplo, han demostrado, a partir del análisis de tareas explicativas, que la dimensión causal o motivacional tendrían mayor preponderancia en relación con las demás.

La representación alcanzada en este nivel indica que la información presentada en el texto se mantendrá en la memoria de largo plazo, debido a las relaciones que establece con el conocimiento ya adquirido por el lector. En este sentido, para Kintsch y Kintsch (2005, p. 76), un lector que no recupera la información desde su memoria de largo plazo con el objetivo de integrarla y relacionarla con la nueva información que está leyendo no comprende un texto, a pesar de que establezca una correcta elaboración del texto base. En la misma línea, Zwaan y Radvansky (1998) señalan que la comprensión del lenguaje significa la construcción de una representación del estado de situaciones descrita en un texto, en otras palabras, el modelo de situación ahí especificado. Con esto, se produjo un cambio radical en la noción de comprensión de lectura con la que se venía operando, al permutar la pregunta *¿cómo los lectores comprenden un texto?* por *¿cómo los lectores construyen un modelo situacional coherente sobre lo que han leído?* Lo anterior, ratifica la noción planteada por Kintsch y Van Dijk (1983) de que los modelos de situación deben constituirse como la base del aprendizaje. Por tanto, en las instituciones educativas, “the goal is often to learn from a text, that is, to construct a situation model that will be remembered and can be used effectively when the information provided by that text is needed in some way in a later time” (Kintsch y Kintsch, 2005, p. 76). Es decir, un uso contextualizado de los textos.

### 2.1.2. REPRESENTACIÓN DE LA COMPRESIÓN A PARTIR DE FUENTES MÚLTIPLES: MODELO DE DOCUMENTOS

Una de las desventajas descrita para los modelos presentados por Kintsch y Van Dijk (1978, 1983) y Kintsch (1988, 1994, 1996, 1998) se relaciona con su incapacidad para operar con los procesos de comprensión derivados de la lectura de múltiples textos. Estas lecturas son las que dominan la mayor parte de las tareas de escritura planteadas en los diferentes niveles educacionales (Rouet, 2006, p. 63), puesto que, como aclara Silvestri (1999, p. 53), “suele observarse en la actualidad una tendencia a descalificar las actividades de aprendizaje a partir de una sola fuente, ya que se asocian con formas de enseñanza tradicionales”.

Diferentes investigaciones se han preocupado de describir los procesos cognitivos involucrados en la comprensión de textos múltiples (Foltz, Britt y Perfetti, 1996; Rouet, Favart, Britt y Perfetti, 1997; Perfetti, Rouet y Britt, 1999; Silvestri, 1999; Cerdán, 2006; Rouet, 2006; Vega, 2011). La lectura a partir de documentos múltiples considera, en cualquier caso, procesos cognitivos que no se manifiestan al momento de leer un solo texto: leer documentos múltiples requiere que los lectores identifiquen el origen de cada documento, comparen la información que en ellos se encuentra y, finalmente, integren la información en una representación coherente (Rouet, 2006, p. 189).

Para Perfetti, Rouet y Britt (1999), emplear de manera adecuada los textos implica mantener la representación mental de cada uno de los textos, la situación en ellos descrita y la relación entre los textos. Se producen representaciones cognitivas que incluyen conexiones, muchas veces implícitas, entre los textos revisados, en tanto los diferentes enfoques entre los textos fuerzan a los lectores a reconocer el vínculo entre los documentos. Según Cerdán (2006, p. 22), los documentos pueden interactuar de diferentes maneras. En primer lugar, pueden sobreponerse; en segundo, un texto puede ser parte del modelo situacional de otro texto; y, por último, los documentos pueden llegar a tener modelos situacionales incompatibles.



El modelo de documentos presentado por Perfetti, Rouet y Britt (1999) consta de dos componentes. Por un lado, el modelo de intertexto representa las relaciones entre los documentos y entre un documento y los elementos de la situación. A su vez, está constituido por nodos de documento y predicados de intertexto. Los primeros se caracterizan por disponer de variables para fuente, metas retóricas y contenido; mientras los segundos representan tanto las relaciones entre documentos, como las relaciones entre un documento específico con un evento del modelo de situación. Por otro lado, el modelo de situación se origina dentro de la concepción presentada en 2.1.1.3. Sin embargo, a diferencia de ella, se considera que textos múltiples establecen múltiples situaciones que el lector debe ser capaz de representar adecuadamente, de acuerdo a los eventos significantes que los documentos establezcan. La representación coherente derivada de la integración de información de los documentos debe implicar la integración de los modelos situacionales en ellos propuestos.

La construcción de una representación integrada de las situaciones se encuentra determinada por el tipo de pregunta o instrucción, en la medida en que el propósito de la lectura a partir de fuentes múltiples es, generalmente, resolver alguna tarea específica: actividad denominada como *razonar con los documentos* (Rouet, Britt, Mason y Perfetti, 1996, cit. en Cerdán, 2006). Perfetti, Rouet y Britt (1999) señalan que existen dos tipos de tareas de escritura originadas en la lectura de múltiples fuentes: por una parte, las que buscan construir un modelo de situación, pero no forzosamente un modelo de intertexto (como los resúmenes); por otra, las que se centran en los documentos con el objetivo de construir un modelo que considere la integración del modelo de intertexto y el modelo situacional derivado de los documentos leídos (como la comparación de documentos). En suma, una representación integrada también debe incluir las relaciones entre los diferentes documentos, con el fin de dar cuenta de las referencias cruzadas, explícitas o implícitas, provenientes de la lectura de los textos (Rouet, 2006, p. 65).

De acuerdo con Silvestri (1999), la realización de una adecuada tarea de fijación<sup>12</sup>, junto con la adecuada representación discursiva de los textos contribuyen, potencialmente, a los aprendizajes auténticos, esto es, a conseguir que el uso de la información adquirida

sea independiente de los requerimientos estipulados. En esta misma línea, Wiley y Voss (1999) indican que existen tareas específicas que promueven la integración y aprendizaje a partir de múltiples fuentes. De hecho, señalan que las actividades que consideren aspectos argumentativos tenían mejores resultados que aquellas que centran su atención en aspectos narrativos o descriptivos de los documentos. Además, éstas se concentraban en el texto en tanto globalidad, con el fin de que las respuestas fueran construcciones coherentes a partir de los documentos. En cambio, la elaboración a partir de preguntas breves responde, por lo general, a otro tipo de objetivos, en tanto insisten sobre el reconocimiento superficial de las ideas de los textos, en otras palabras, fijan su atención en tareas cognitivas –derivadas de la lectura de textos– menos complejas, como la extracción de información. Por tanto, para Cerdán (2006, p. 40), la atención de los estudiantes no se concentra en la integración de la información, sino que, simplemente, prestan especial cuidado en extraer las unidades informativas requeridas.

## 2.2. ANÁLISIS SEMÁNTICO LATENTE

### 2.2.1. EL ANÁLISIS SEMÁNTICO LATENTE Y SU RELACIÓN CON EL MODELO DE KINTSCH

La investigación sobre los niveles de representación del discurso ha estado dominada por los análisis de las proposiciones que conforman un determinado texto. A pesar de lo útil que ha resultado el enfoque, implica una carencia operacional no menos significativa, en tanto el proceso, difícilmente, se puede automatizar. Dado lo anterior, la proposicionalización de los documentos sería factible solo si estos fueran de pequeña extensión, por lo que operar con largas colecciones de textos resultaría imposible (Kintsch, 1998, p. 86). Por otro lado, considerar la información presente en los textos de la misma manera en que se trata el *conocimiento* permite que la *red de conocimiento* se establezca con escaso orden, lo que supone una estructura que no está creada con anterioridad, sino que se activa en el contexto de la operación requerida (Kintsch, 1996, p. 76). El significado de los nodos constituyentes de la *red de conocimiento*, por tanto, es “específico de la

situación y depende del contexto” (Kintsch, 1996, p. 78); vale decir, está determinado por los valores de cercanía entre los nodos.

Los conceptos o proposiciones se consideran no como nodos formantes de una red, sino como vectores en un espacio semántico multi-dimensional, realizado como constituyente del análisis con LSA. Al respecto, Kintsch (1998, p. 87) asegura: “A concept or proposition can thus be thought of as a vector of numbers, each number indicating the strength with which the concept or proposition is linked to another concept or proposition”. Con esta representación vectorial, se supera la subjetividad que dominaba el análisis proposicional de los documentos, en la medida en que la consideración de los conceptos como vectores en un espacio semántico es automática y objetiva, al relacionar los nodos a través de una medida de distancia *coseno* (ver apartado 2.2.2.) y no arbitrariamente, al depender de cada investigador (Kintsch, 1998, p. 89).

LSA opera con el recuento de una gran cantidad de palabras que ocurren en contextos similares. Sin embargo, a diferencia de las proposiciones, las palabras no se consideran como “unidades apropiadas de la cognición” (Kintsch, 1998, p. 88). Para evitar definir las palabras en términos de documentos y los documentos en términos de palabras, el LSA sustituye esta ambigüedad gracias a una aproximación semántica, que elimina gran parte de información innecesaria, por lo que se mantiene, solamente, su estructura semántica esencial (denominada *descomposición en valores singulares*, ver apartado 2.2.2.), con resultados similares a los que se obtendrían con un análisis proposicional.

Por ejemplo, Kintsch (2007) precisa que la mejor manera de representar el significado que tienen las palabras en la mente es como vectores en un espacio semántico. De esta forma, según LSA, el significado de las lexías no estaría definido (como ocurre en los diccionarios), sino que se determinaría a partir de su ubicación en relación con todas las demás palabras del espacio semántico. Por tanto:

Meaning, for LSA, is a relation among words. In such a relational system, one cannot talk about the meaning of a word isolation; words have meaning only by virtue of their relations to other words –meaning is a property of the system as a whole (p. 91).

Esta concepción automatizada del análisis de la representación de la comprensión del discurso se relaciona directamente con la propia definición del concepto presentada por Kintsch (1996), esto es, elaborar una representación de lo comprendido sobre la cual puedan realizarse múltiples cálculos. Para caracterizar automáticamente la comprensión lectora de las personas a través de LSA, vale decir, la construcción adecuada de un modelo situacional a partir de la teoría de Kintsch, se han propuesto diversos estudios (Lemaire, Denhière, Bellissens y Jhean-Larose, 2006; Lemaire, Denhière, Bellissens y Jhean-Larose, 2007; Rus, Lintean y Azevedo, 2009) que se han orientado hacia la descripción del modelo situacional desde la comprensión un único documento. Sin embargo, los esfuerzos por caracterizar la comprensión a partir de múltiples documentos con LSA ya ha sido trabajado por Foltz, Britt y Perfetti (1996) quienes compararon la calidad de la comprensión de lectores novatos y expertos sobre la base de una gran cantidad de ensayos de *Historia*. Los autores concluyen que tanto expertos como novatos incluyen información derivada de su conocimiento previo que afecta en los resultados del análisis semántico latente. Por otra parte, además, se han postulado más características (denominadas *Latent Semantic Features*), que permitirían identificar el modelo situacional alcanzado por una persona a partir de la lectura de dos documentos (Bravo-Márquez, L'Huillier, Moya, Ríos, Velásquez, 2011). De esta manera, las distancias LSA establecidas más el empleo de ciertos tipos de conectores textuales ayudarían a comprobar el nivel de comprensión del discurso.

### 2.2.2. IMPLEMENTACIÓN DEL LSA

LSA se concibe como una teoría computacional para la adquisición y representación del conocimiento y se define como “a fully automatic mathematical/statistical technique for extracting and inferring relations of expected contextual usage of words in passages of discourse” (Landauer, Foltz y Laham, 1998, p. 8).

El LSA puede ser comprendido de dos maneras diferentes. Por un lado, como una forma práctica para obtener resultados estimados aproximados –según distancia *coseno*<sup>13</sup> – del uso contextual de una palabra al interior de una gran colección de documentos, lo que

evidencia una alta correlación entre los resultados arrojados por LSA y las representaciones de las personas sobre lo que han escuchado o leído (Landauer, Foltz y Laham, 1998, p. 4). Por otro, como un modelo de procesos computacionales y representaciones implicados en la adquisición y utilización del conocimiento.

Con el fin de construir el modelo del espacio vectorial, indispensable para el LSA, se deben realizar dos pasos (cf. Landauer, Foltz y Laham, 1998; Venegas, 2005; Martin y Berry, 2007). Primero, se desarrolla la matriz *tipo por documento* (ver Cuadro 5 y Cuadro 6) con los *pesos* de cada una de las palabras de los textos, esto es, las palabras que presentan una baja ocurrencia tienen un peso alto, en tanto a las de mayor frecuencia se les asigna un peso más bajo (Cuadro 7) (Gutiérrez, 2005). En la matriz *tipo por documento*, las filas corresponden a los tipos –también denominados *signos tipos*, como parte de la dicotomía tipo/ocurrencia (*type/token*<sup>14</sup>)– que conforman un documento. Si bien lo más común es que los tipos correspondan a términos, pueden estar constituidas por ocurrencias, frases o conceptos, de acuerdo a lo que se requiera investigar. Las columnas, por su parte, representan los documentos en los que se encuentran los términos. En consecuencia, la colección de los documentos se compone de  $n$  documentos y  $m$  términos que puede ser representado como  $m$  por  $n$  tipo por documento de la matriz.

Cuadro 5: *Ejemplo de títulos de tópicos sobre música y cocina*

<b>Documento</b>	<b>Título</b>
<b>1</b>	Rock and Roll Music in the 1960's
<b>2</b>	Different Drum Rolls, a Demonstration of Techniques
<b>3</b>	Drum and Bass Composition
<b>4</b>	A perspective of Rock Music in the 90's
<b>5</b>	Music and Composition of Popular Bands
<b>6</b>	How to make Bread and Rolls, a Demonstration
<b>7</b>	Ingredients for Crescent Rolls
<b>8</b>	A Recipe for Sourdough Bread
<b>9</b>	A Quick Recipe for Pizza Dough using Organic Ingredients

*Nota 1.* Extraído de Martin y Berry (2007, p 37).

Cuadro 6: *Matriz tipo por documento en relación con los datos del Cuadro 5*

<b>Tipos (m)</b>	<b>Documentos (n)</b>								
	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>
<b>Bread</b>	0	0	0	0	0	1	0	1	0
<b>Composition</b>	0	0	1	0	1	0	0	0	0
<b>Demonstration</b>	0	1	0	0	0	1	0	0	0
<b>Dough</b>	0	0	0	0	0	0	0	1	1
<b>Drum</b>	0	1	1	0	0	0	0	0	0
<b>Ingredients</b>	0	0	0	0	0	0	1	0	1
<b>Music</b>	1	0	0	1	1	0	0	0	0
<b>Recipe</b>	0	0	0	0	0	0	0	1	1
<b>Rock</b>	1	0	0	1	0	0	0	0	0
<b>Roll</b>	1	1	0	0	0	1	1	0	0

*Nota 1.* Extraído de Martin y Berry (2007, p. 38).

Cuadro 7: Pesos de la matriz tipo por documento en relación con los datos del Cuadro 6

Tipos ( <i>m</i> )	Documentos ( <i>n</i> )								
	1	2	3	4	5	6	7	8	9
<b>Bread</b>	0	0	0	0	0	.474	0	.474	0
<b>Composition</b>	0	0	.474	0	.474	0	0	0	0
<b>Demonstration</b>	0	.474	0	0	0	.474	0	0	0
<b>Dough</b>	0	0	0	0	0	0	0	.474	.474
<b>Drum</b>	0	.474	.474	0	0	0	0	0	0
<b>Ingredients</b>	0	0	0	0	0	0	.474	0	.474
<b>Music</b>	.347	0	0	.347	.347	0	0	0	0
<b>Recipe</b>	0	0	0	0	0	0	0	.474	.474
<b>Rock</b>	.474	0	0	.474	0	0	0	0	0
<b>Roll</b>	.256	.256	0	0	0	.256	.256	0	0

Nota 1. Basado en Martin y Berry (2007, p. 39).

Luego de esta construcción, la matriz se convierte en un espacio vectorial de tipo por documento, a través de una descomposición ortogonal, que permite mantener propiedades esenciales presentes en la matriz original. La matriz ortogonal se define a partir de la expresión  $Q^T Q = I$ , donde  $Q$  es una matriz ortogonal,  $Q^T$  es la matriz  $Q$  transpuesta e  $I$  es la matriz original, tal como se detalla en la Figura 3.

Figura 3: Representación de la matriz ortogonal<sup>15</sup>

$$Q^T Q = \begin{bmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & & 0 \\ \vdots & 0 & \ddots & \ddots & \vdots \\ 0 & & \ddots & 1 & 0 \\ 0 & 0 & \dots & 0 & 1 \end{bmatrix}$$

De acuerdo con Martin y Berry (2007, pp. 40-41), existen varios métodos para realizar la descomposición de la matriz. La estrategia más empleada es la *descomposición*

*en valores singulares* (SVD, por su sigla en inglés), por varios motivos. En primer lugar, la SVD descompone la matriz en factores individuales tanto para términos como para documentos, que se consiguen simultáneamente. En segundo lugar, distingue la estructura semántica subyacente a la colección de documentos y, además, permite ajustar la matriz gracias a la reducción de dimensiones. En tercer lugar, la SVD es capaz de operar con una gran cantidad de datos.

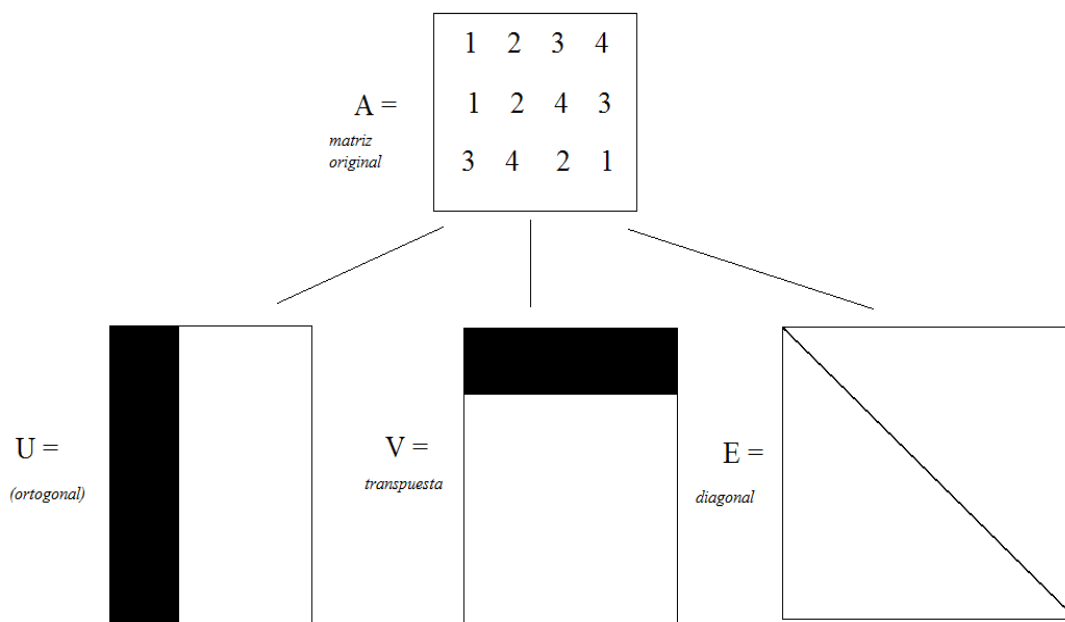
La SVD descompone la información de la matriz *tipo por documentos* y la representa en otras 3 matrices características inherentes (singulares) de los documentos. La reducción se realiza a partir de la eliminación de aquellas palabras con un peso inferior, permitiendo la normalización de las distribuciones de frecuencia de los documentos; su resultado se identifica con un espacio semántico compacto, por tanto, sencillo de analizar de manera automatizada (Gutiérrez, 2005).

De esta manera, la matriz (A) (ver Figura 4) se especifica a partir de la siguiente expresión  $A=U\Sigma V^T$ , donde:

- a. U es una matriz ortogonal ( $U^T U = I_m$ ), lograda a partir del procesamiento lineal del número total de columnas de la matriz original;
- b. V es una matriz ortogonal ( $V^T V = I_n$ ), lograda a partir de la permutación de filas por columnas; y
- c.  $\Sigma$  es una matriz diagonal ( $\Sigma = \text{diagonal}(\sigma_1, \sigma_2, \dots, \sigma_n)$ ), lograda a partir del procesamiento lineal del número de filas, del número de columnas y la cantidad de dimensiones de la matriz primaria (A) (Gutiérrez, 2005).



Figura 4: Representación de SVD<sup>16</sup>



### 2.3. ESCRITURA ACADÉMICA

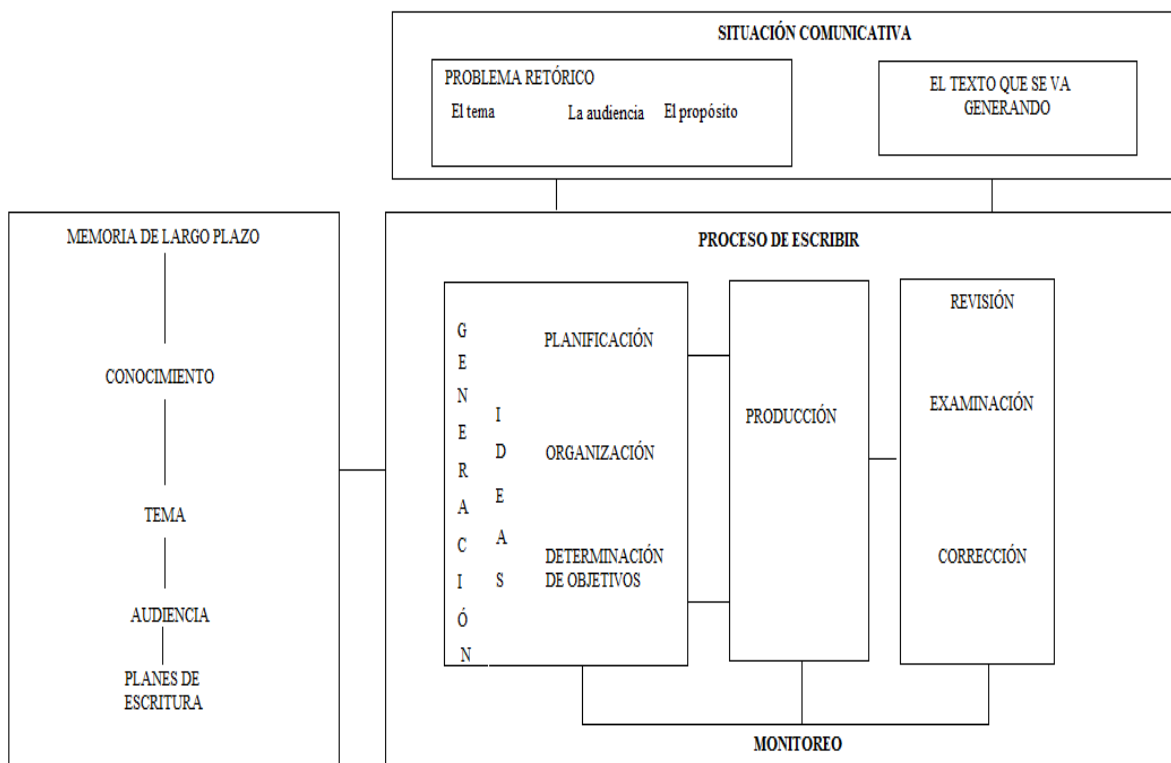
Por lo general, para evaluar la representación de la comprensión del discurso, los alumnos deben elaborar textos escritos a partir de instrucciones dadas por los profesores (cf. Solé, s.f.). Además, para los propósitos de la presente investigación, una de las características de la comprensión de documentos se relaciona con la capacidad que tienen las personas para integrar, en un solo texto, información relevante proveniente de todas las fuentes consideradas (Rouet, 2006). En este sentido, dado que se opera con textos producidos en un contexto académico, el trabajo se inscribe, asimismo, dentro de la escritura académica.

El estudio de la escritura en el nivel académico o científico presenta tres grandes aproximaciones teóricas (Hyland, 2002). La primera de ellas se centra en el análisis de las estructuras lingüísticas o retóricas utilizadas por los escritores para elaborar sus textos. Concretamente, de acuerdo con los fines adscritos a esta corriente, la investigación se

puede focalizar tanto en la descripción de la forma en que se organizan las regularidades microestructurales de los documentos (es decir, define la conciencia que tiene el escritor sobre el sistema de reglas adecuado para la construcción de textos), como en la caracterización de las estructuras retóricas que le permitirían al escritor comunicarse con sus lectores. Con todo, si bien este movimiento cuenta con numerosos trabajos, se le critica su postura contraria a la influencia social que cumplirían los textos, lo que posibilitaría explicar su naturaleza a partir, exclusivamente, de su análisis intrínseco (cf. Bazerman y Paradis, 1991, p. 3).

La segunda tendencia considera que los esfuerzos investigativos no deben enfocarse en el texto, sino que en el escritor; en otras palabras, se orienta, principalmente, a la descripción de los procesos cognitivos que están involucrados en el proceso de escritura. En esta perspectiva, se erigen dos posturas como fundamentales. Por un lado, los aportes realizados por Flowers y Hayes (1981, cit. en Hyland, 2002) apuntan que el proceso de escritura se encuentra determinado por la tarea de escritura y por la memoria de largo plazo. Todas las características del modelo (ver Figura 5) se evidenciarían en los escritores expertos, al contrario de lo que ocurriría con aquellos que no lo son. Por otro lado, el modelo de Bereiter y Scardamalia (1987, cit. en Hyland, 2002) señala que la distancia entre expertos y novatos sería tan evidente que se requieren dos modelos diferentes para definirlos. Mientras el modelo de *decir el conocimiento* representa a aquellos escritores que se limitan a decir lo que recuerdan a partir de la tarea, el modelo de *transformar el conocimiento* supone que las habilidades de los escritores son tales que les permiten reflexionar acerca de las complejidades de la tarea y resolver problemas de contenido, forma, audiencia, etc. Pese a que esta concepción de la escritura como un proceso que consta de ciertas etapas ha tenido una fuerte influencia en las instituciones de enseñanza, Hyland (2002, p. 29) advierte que existe poca evidencia que demuestre una mejora en las habilidades de escritura, al mantener este modelo de instrucción.

Figura 5: *Modelo de Flowers y Hayes (1981)*<sup>17</sup>



Por último, el tercer enfoque reconoce que la escritura constituye tanto un esfuerzo cognitivo como interactivo. En éste, uno de los aportes más significativos deriva de los estudios que entienden a la escritura como un proceso de construcción social, que estructura las relaciones entre las personas y organiza la percepción sobre el mundo (Bazerman y Paradis, 1991, p. 3). Para lograrlo, el escritor se adscribe a una determinada *comunidad discursiva*, compuesta por: “conventions, a vital history, mechanisms for wielding power, institutional hierarchies, vested interests and so on” (Porter, 1992, cit. en Hyland, 2002, p. 41). En otras palabras, el escritor no se encuentra limitado a escribirle a un determinado lector, sino que debe ceñirse a las características impuestas desde su comunidad, con el objetivo de comunicarse de manera efectiva. Su voz siempre está relacionada con las ideas de los demás miembros de la comunidad, lo que permite, a su vez, el establecimiento de la identidad particular y la legitimación de su discurso. Estas comunidades discursivas pertenecen a determinadas disciplinas, que poseen características particulares para la

transmisión de información: normas, nomenclatura, cuerpos de conocimiento, conjunto de convenciones y modos de investigar, entre otros (Hyland, 2000, p. 8). La diferencia entre las disciplinas no se acaba tan solo en *lo* que transmiten, sino en *cómo* lo transmiten.

Un número importante de estas investigaciones se ha centrado en describir las características discursivas disciplinarias a partir de los géneros textuales prototípicos que mueven el saber disciplinario. Sin embargo, la función de la escritura en los primeros niveles universitarios no se relaciona, necesariamente, con la apropiación de los rasgos que definen a los géneros discursivos propios de la disciplina, sino que, más bien, se centran en evaluar contenidos, independientemente del género empleado. Al respecto, Moyano (2010, p. 467) afirma que en todos los estudios sobre escritura académica enfocados en el nivel lingüístico textual o en el nivel de la caracterización de los tipos textuales “se ha hecho una escasa o ninguna mención a la enseñanza, desde una perspectiva funcional, de los usos específicos de recursos lingüísticos y discursivos para la construcción de significado en los diferentes grupos de disciplinas”.

Los estudiantes de la enseñanza superior (sobre todo aquellos que recién ingresan) presentan una serie de dificultades al momento de comunicarse por escrito. No solo se manifiestan problemas con la microestructura de los textos, relacionados con la escasa experiencia de los estudiantes en transformar el conocimiento, sino que, principalmente, los alumnos se muestran incapaces de adquirir todos los requerimientos impuestos desde su nueva comunidad de comunicación rápidamente. A este proceso se le ha denominado *alfabetización académica*. Carlino (2005, pp.13-14) distingue al respecto dos significados, uno sincrónico, definido como: “el conjunto de nociones y estrategias necesarias para participar en la cultura discursiva de las disciplinas así como en las actividades de producción y análisis de textos requeridas para aprender en la universidad”; y otro diacrónico, que especifica el proceso gracias al cual se ingresa a la comunidad discursiva. En consecuencia, al ser comprendido como un proceso, no se concibe que las habilidades escriturales sean adquiridas para siempre, por el contrario, dada la gran diversidad de asignaturas, tareas, motivaciones, etc., el ingreso a las culturas disciplinarias no se logra por

completo, ni siquiera, al concluir los estudios superiores (cf. Carlino, 2005; Berkenkotter, Huckin y Ackerman, 1991).

## 2.4. DISCURSO AJENO

### 2.4.1. TRADICIÓN GRAMATICAL

Gran parte de los estudios lingüísticos relacionados con la descripción del tratamiento de la información ajena se encuentra en diversos estudios gramaticales (Real Academia de la Lengua, 1973 y 2010; Maldonado, 1999; Reyes, 1995). A pesar de que el tratamiento se concentra, principalmente, en las condicionantes sintácticas que permiten la aparición de los discursos ajenos, difieren en la denominación misma del fenómeno. En primer lugar, la Real Academia de la Lengua (2010, p. 833-834) integra el concepto de *Discurso directo* (DD) y *Discurso indirecto* (DI): mientras el primero “reproduce de forma literal palabras o pensamientos”, en el segundo las palabras ajenas se adaptan al sistema referencial deíctico de quien se encuentra creando el discurso.

En segundo lugar, Maldonado (1999) escoge el concepto de ‘discurso reproducido’, en lugar del ‘discurso referido’. Mientras “este último tan solo describe una acción realizada verbalmente, el primero reproduce una situación de enunciación (reproduce, por tanto, cuál fue el enunciado original, quiénes fueron el hablante y el destinatario del mensaje)” (p. 3556). De esta manera, entiende el ‘discurso reproducido’ como un universal lingüístico que se puede manifestar de diferentes maneras, tal como se muestra en el Cuadro 8.

Cuadro 8: *Gradación en la presentación del discurso reproducido*

Tipo	Ejemplo
Mención del suceso de habla	Anoche estuve charlando con un amigo
Descripción general de lo hablado	Me felicitó efusivamente
Resumen del contenido	Me dijo que estaba orgulloso de mí, que se había alegrado mucho por mi éxito, y todas esas cosas que se dicen en estos casos
Cita literal de las palabras ajenas	Me dijo emocionado: “¡Enhorabuena!”
Monólogo interior en primera persona	¡Qué ilusión me hizo! Fue muy majo al felicitarme. Porque yo sé que, en el fondo, tenía pelusa. ¡Si lo sabré yo! Había reaccionado bien, pero no sé hasta qué punto esa reacción era sincera.

*Nota 1.* Basado en Maldonado (1999, p. 3551).

Dentro de esta gradación, afirma que tanto el DD como el DI se pueden manifestar de diferentes maneras, entre las que destacan el ‘discurso directo libre’, que carece de un verbo que introduzca la cita; el ‘discurso pseudo-directo’, en donde se resume el contenido y se agregan fragmentos entre comillas; el ‘discurso indirecto mimético’, que admite componentes agramaticales que serían aceptables solamente en el DD. Para distinguirlos, señala que el DD se establece a partir de una ‘expresión introductora’, que incluye un verbo de ‘decir’ conjugado y una ‘cita directa’, identificable por alguna marca tipográfica (guión o comillas). Por otro lado, el DI implica una ‘expresión introductora’, con un verbo de ‘decir’ conjugado y una ‘cita indirecta’ que se relaciona con el verbo de la expresión introductora por medio de la conjunción subordinante *que* (1999, p. 3554). De lo anterior, se desprende que la autora realiza la distinción entre el discurso del hablante y el discurso reproducido, al igualar este último con las citas y ponerlas en función de los requisitos (y preferencias) sintácticas del hablante, a partir de las condiciones necesarias establecidas para su correcta aparición (ver Cuadro 9).

Cuadro 9: *Condiciones caracterizadoras del discurso reproducido*

<b>Condición</b>	<b>Definición</b>	<b>Ejemplo</b>
Primera	El objeto de una situación de enunciación determinada debe ser otra situación de enunciación que se reproducirá	Mi madre me dijo que mi tía le había dicho que sí. *Mi madre me dijo: Tu tía me ha dicho: Mis hijos me dicen: Tu hija ha dicho <sup>18</sup>
Segunda	El objeto del discurso se debe representar en la cadena verbal que se refiere a la situación de enunciación del hablante	Me gritó que me marchara. *Gritó.
Tercera	El verbo de la expresión introductora no puede utilizarse con valor realizativo	Te lo avisé: Te han descubierto. *Te lo aviso: Te han descubierto.

*Nota 1.* Basado en Maldonado (1999, p. 3555-3557)

Por último, Reyes (1995) prefiere los términos estilo directo (ED) y estilo indirecto (EI), para centrarse tanto en la descripción de las características sintácticas como pragmáticas de estas realizaciones. Sin embargo, a diferencia de Maldonado (1999), trata indistintamente los conceptos de ‘cita’ y ‘estilo’, aunque es consciente de que se debe diferenciar entre “la voz que cita y la voz citada” (p.9). En este contexto, señala que citar implica que alguien atribuye, de manera intencional a otro hablante, cierto fragmento de discurso.

Para Reyes (1995) tanto el ED como el EI son los recursos más utilizados para citar en todos los ámbitos. La diferencia entre ellos se establece a partir de la forma en que se reproducen las palabras ajenas; mientras, en el ED se mantienen idénticas, en el EI presentan variaciones, en tanto el hablante las adecúa a su propia situación de enunciación. Esta adaptación del EI no se presenta de una sola manera, sino que se puede manifestar como una reconstrucción, en ocasiones muy libre, de un discurso (Reyes, 1995, p. 13). En

cuanto funciones, se plantea que las palabras utilizadas en el EI implican que el hablante, por un lado, desea transmitir contenidos factuales (Reyes, 1995, p. 45) más que afectivos (sin embargo, también se pueden verificar contenidos cognitivos no factuales, como con el verbo *pensar*) y que, por otro, ha interpretado las palabras ajenas antes de reproducirlas en su propio acto de enunciación.

#### 2.4.2. EL DISCURSO AJENO EN LA ESCRITURA ACADÉMICA

La integración de la información en la escritura académica se supone como un proceso difícil con patrones complejos de aparición (Soto, 2009). En este sentido, cada texto creado dentro de cualquier comunidad discursiva se relaciona con los demás que participan en ella, a través del proceso denominado *intertextualidad*. El término, integrado ya por Bajtín (1982, cit. en Hyland, 2002, p. 36), se caracteriza a partir de la noción de que todos los discursos se relacionan con otros discursos, en la medida en que su construcción está mediada por una serie de textos, que forman parte del conocimiento de los hablantes. Para De Beaugrande y Dressler (1981), la intertextualidad se refiere a los factores que hacen depender la utilización adecuada del conocimiento de un texto en relación con otros textos anteriores (p. 45). Con respecto a lo anterior, ningún discurso se refiere a aspectos que antes no hayan sido dichos, por tanto, la originalidad de los textos radica, esencialmente, más en la forma en que cada autor relaciona sus ideas con las demás, que en la novedad de lo planteado (cf. Bazerman, 2004, p. 83). De acuerdo con Bazerman (2004, pp.88-89), se identifican seis técnicas para la representación de la intertextualidad, detalladas en el Cuadro 10. Para nuestros fines investigativos, los más utilizados son los primeros dos (cita directa e indirecta), en la medida en que son los más fácilmente reconocibles. Además, su importancia radica en que las citas, específicamente, para Pecorari (2008, p. 53), atribuyen cierta proposición a otra fuente, pero cuando se carece de dicha atribución, el escritor asume la proposición como una afirmación propia.



Cuadro 10: *Técnicas de representación intertextual*

<b>Tipo</b>	<b>Caracterización</b>
Cita directa	Identificada a partir de marcas tipográficas, como por ejemplo cursivas.
Cita indirecta	Usualmente, especifica una fuente y se aproxima a reproducir el significado original, pero en palabras del escritor que hace uso de ella.
Mención de una persona o documento	Mera mención de un documento, basada en la familiaridad del lector con el tema
Comentario de un fragmento de texto	Significa que a través de la evaluación sobre algún aspecto se reconoce implícitamente a otra voz
Uso de una terminología reconocible asociada a un determinado grupo	Empleo de términos específicos de grupo
Uso de lenguaje y formas que representan ciertos tipos de comunicación	Empleo de tipos de vocabulario o registro, frases hechas que implican pertenencia a un género

*Nota 1.* Basado en Bazerman (2004, pp. 88-89).

Una de las prácticas más importantes en la escritura académica se relaciona con la interacción con la audiencia por medio de la atribución de la información, en otras palabras, citar la información ajena es una de las estrategias que permiten transformar y validar las creencias en conocimiento. Además, la correcta atribución se constituye como una habilidad esencial para la aceptación de los postulados presentados, tal como señala Hyland (2000, p. 21):

Appropriate textual practices are vital to acceptance of claims. Explicit reference to prior literature is a substantial indication of a text's dependence on contextual knowledge and thus a vital piece in the collaborative construction of new knowledge between writers and readers. The embedding of arguments in networks of references not only suggests and cumulative and linear progression, but reminds us that statements are invariable a

response to previous statements and are themselves available for further statements by others.

Se constituye, por tanto, como un proceso social en el que el escritor debe negociar su posición dentro de la comunidad académica, gracias a la contextualización de su obra en el universo general de las investigaciones relacionadas. De esta manera, se considera que el uso de las citas es un elemento primordial de los artículos modernos de investigación, puesto que a través de ellas, los escritores construyen los hechos en sus comunidades discursivas (Hyland, 2000).

Al existir diferentes culturas disciplinarias que imponen sus propias características textuales, las formas en que se integra la información desde otra fuente demuestran notables diferencias, en tanto sus regularidades apuntan a acciones sociales estereotípicas de las comunidades (Hyland, 2000, p. 29). Estas diferencias se han analizado, principalmente, a partir de la dicotomía analítica planteada por Swales (1990), entre citas integrales y no integrales. Mientras las primeras se identifican gracias a que el nombre del investigador se menciona en la misma oración y forma parte de ella, en las citas no integrales el nombre del investigador aparece entre paréntesis o es referido mediante otro mecanismo (como nota al pie de página). Desde esta perspectiva, Hyland (1999) analizó 80 artículos de investigación publicados en revistas especializadas y concluyó que las comunidades que forman parte del *soft knowledge* –como Filosofía– utilizan más las citas integrales que las que son parte del *hard knowledge* –como Física (ver Cuadro 11).

Cuadro 11: *Porcentaje de tipos de citas empleados en las distintas disciplinas*

<b>Disciplina</b>	<b>No integral</b>	<b>Integral</b>
<b>Biología</b>	90,2	9,8
<b>Ingeniería electrónica</b>	84,3	15,7
<b>Física</b>	83,1	16,9
<b>Ingeniería en matemática</b>	71,3	28,7
<b>Marketing</b>	70,3	29,7
<b>Lingüística aplicada</b>	65,6	34,4
<b>Sociología</b>	64,6	35,4
<b>Filosofía</b>	35,4	64,6

*Nota 1.* Extraído de Hyland (1999)

#### 2.4.3. PLAGIO ACADÉMICO

En Chile, un estudio para determinar las percepciones y prácticas de estudiantes en torno al plagio realizado a 2.013 estudiantes de educación media de la región Metropolitana y de Valparaíso, y 1.126 estudiantes de educación superior solo de la región Metropolitana<sup>19</sup> concluyó que un 79,6% de los estudiantes secundarios copió y pegó información de la *Web*<sup>20</sup> sin indicar la fuente, al menos una vez durante el año 2009; mientras el 64,1% de los estudiantes universitarios lo hizo en el mismo periodo (Molina, et al., 2011). La explicación a dicha conducta, tal como se detalla en el Anexo 1, se debe a que, principalmente, les permite ahorrar tiempo y porque creen que el trabajo será de una mejor calidad que si lo hubiesen escrito ellos mismos. A pesar de que estos datos son muy escasos para poder emitir algún juicio contundente, es posible señalar que, efectivamente, dichos resultados se correlacionan con las respuestas que dan estudiantes en Europa (cf. Pupovac, Bilič-Zulle, Petrovečki, 2008), lo que podría indicar una tendencia global sobre el tema.

Sin duda, el plagio se ha investigado con mayor intensidad en los últimos años por la influencia que ha tenido la *Web* sobre las formas de acceso la información. La presencia

de la *Web* no tan solo ha implicado una evolución en las prácticas de búsqueda de contenidos adecuados desde un punto de vista académico, sino que, asimismo, de acuerdo con Sutherland-Smith (2008, p. 100), ha variado la noción de autoría y propiedad de los textos que se presentan en un ambiente hipertextual. Es más, la *Web*, junto con el uso de procesadores para la elaboración de los documentos, facilitaría la construcción de textos por medio de la copia de grandes fragmentos. Por otro lado, la autora precisa que coexisten investigaciones que apoyan tanto la influencia directa que existiría entre el uso de la *Web* y el plagio, como las que no. En esta última postura, Moore Howard (2007, p. 4) critica la idea de que la *Web* incremente la presencia de plagio, puesto que no aborda las complejidades textuales que han adoptado los textos en entornos hipermediales. Apunta a la necesidad de que el fenómeno del acceso y uso de la información en estos medios sea parte fundamental de una educación sobre las nuevas tecnologías, con el objetivo de superar las barreras y miedos culturales asociados a la extracción de contenidos como un fenómeno, también, de intertextualidad.

Con el fin de detener el plagio, se han diseñado diversas herramientas computacionales que ayudan a determinar la procedencia de los fragmentos sospechosos de haber sido copiados (cf. Lancaster, 2004; Češka, 2009; Guerrero, 2010). La categorización de estas herramientas descansa según la complejidad del método empleado y en el número de documentos que se procesan para determinar su similitud (ver Cuadro 12).

Cuadro 12: *Categorización de los métodos de identificación de plagio*

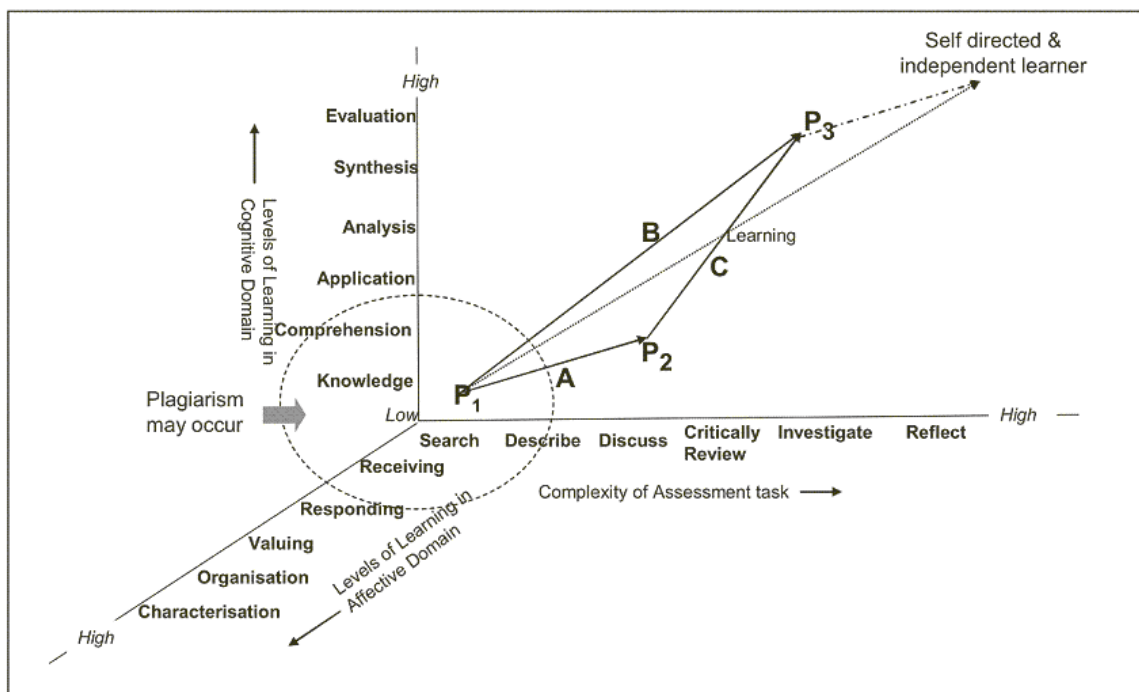
<b>Tipo de clasificación</b>	<b>Descripción</b>	
<b>Complejidad del método utilizado</b>	Superficial	La métrica de similitud se calcula sin ningún conocimiento de las normas lingüísticas, ni una estructura de documento.
	Estructural	La métrica de similitud se calcula con una comprensión parcial de los documentos.
<b>Cantidad de documentos procesados</b>	Singular	Se procesa un solo documento para calcular las métricas de similitud, para lo cual deben evaluarse dos parámetros independientes.
	Pares	Dos documentos se procesan para extraer las métricas de similitud.
	Multidimensional	$N$ documentos de un corpus se procesan para determinar las métricas de similitud entre ellos.
	Corpal	Todos los documentos del corpus son procesados para extraer las métricas de similitud.

*Nota 1.* Adaptado de Češka (2009, p. 12).

Sin embargo, Bhattacharya y Jorgensen (2008, p. 204) proponen que una de las estrategias con mejores resultados en la reducción del plagio académico se relaciona con los objetivos que deben buscar los profesores al momento de solicitar trabajos. Afirman que si bien no se puede (ni debe) detener el uso de la *Web* como fuente de información a los estudiantes, sí es posible controlar lo que se desea hacer con dicha información. En definitiva, en cualquier caso, las investigaciones, tareas, pruebas, entre otro tipo de

evaluaciones, tienen que estar orientadas al aprendizaje y no a la repetición textual de los contenidos. Al lograrse, se disminuirían los casos de plagio (ver Figura 6).

Figura 6: Gráfico de representación del objetivo de la tarea en relación con el plagio<sup>21</sup>



Una de las grandes dificultades que enfrentan los estudiantes al momento de escribir un texto es determinar la técnica de representación intertextual más adecuada para sus propósitos (Cuadro 10). La utilización de citas directas o indirectas (que implican la mención de la procedencia de la información) dependería del grado de conocimiento de la información por parte de la comunidad discursiva, es decir, desde qué momento la información que se está entregando debe contar con una fuente identificable para los propósitos buscados. Diversas instituciones de educación superior –en especial de Europa y Norteamérica– han facilitado instructivos explicando lo que debe ser entendido como *conocimiento común* (‘Common Knowledge’) (Haviland y Mullin, 2009, pp. 180-184), en la medida en que comprenden que este concepto determinaría el límite entre lo citable y lo no citable. No obstante, la noción de *conocimiento común* carece de un significado

transversal, en tanto las definiciones se centran en diferentes aspectos que impiden su masificación y utilización. Aun así, la mayoría se basa en las características discursivas de las disciplinas, tal como se desprende de la definición de Hairston y Ruszkiewicz (1993, cit. en Haviland y Mullin, 2009, p. 184) y de la pauta para evitar esta falencia entregada por la Universidad de Oregon (cit. en Haviland y Mullin, 2009, p. 184):

‘facts, dates, events, information, and concepts that belong generally to an educated public. No individual owns the facts about history, physics, social behavior, geography, current events, popular culture, and so on.’ Therefore, common knowledge does not need to be cited—the difficulty is knowing when something is, in fact, widely known. An added twist is that each discipline has its own common knowledge, for example, psychologists will be familiar with the work of Jean Piaget so you do not need to establish who he was.

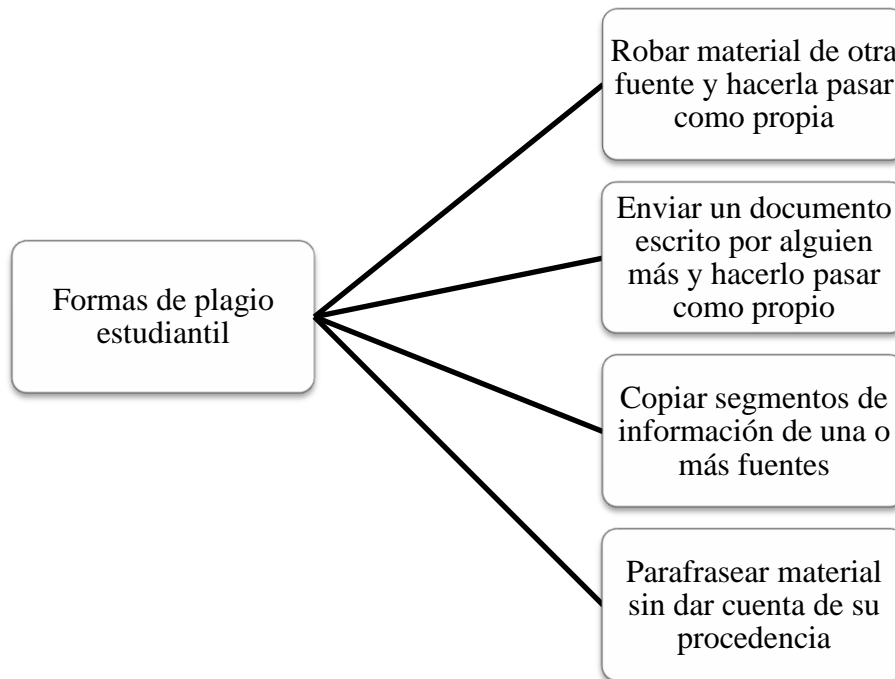
En consecuencia, la cantidad y el tipo de información que contenga referencias explícitas acerca de su procedencia dependerán de la propia disciplina. Con todo, los alumnos al momento de ingresar (o incluso en años intermedios de su formación) desconocerían cuál es la información compartida por los miembros de la comunidad y que, por tanto, no debe ser citada de forma obligatoria. En este sentido, se ha planteado que los profesores deberían conducir este aprendizaje en los años iniciales de la enseñanza universitaria (cf. Haviland y Mullin, 2009; Sutherland-Smith, 2008).

Se identifican, principalmente, cuatro maneras en la que los estudiantes plagian documentos (ver Figura 7). Sin embargo, los límites textuales, considerando la cantidad de palabras que implica un *plagio* no son claros, como sí lo son en el *plagio literario* a partir de la traducción de obras. La *Lingüística forense* ha podido establecer criterios cuantitativos que permiten determinar este tipo de plagio a base del porcentaje de *vocabulario coincidente* entre dos documentos. Al respecto, Turell (2005, p.281) señala:

una medida significativa que permite establecer el nivel de similitud entre varios textos es el *vocabulario coincidente*. Es bastante normal que el parecido entre 2 textos de un

mismo tema oscile entre el 30% y el 40% y que cualquier parecido por encima del 50%, apuntaría hacia la existencia de algún tipo de relación entre los textos analizados.

Figura 7: *Esquema de las cuatro formas en que los estudiantes cometen plagios*<sup>22</sup>



En consecuencia, según Pecorari (2008), el plagio académico debe ser revisado desde un criterio lingüístico para que su análisis resulte objetivable y no se centre en subjetividades inherentes al uso incorrecto de la información, en la medida en que influyen otros factores más allá de la mera repetición léxica coincidente. Una primera aproximación a las características lingüísticas del plagio académico es recogida por Sutherland-Smith (2008, basada en Pecorari, 2002), quien sugiere que está compuesto por 6 rasgos esenciales:

- i. un objeto de lenguaje;
- ii. que ha sido tomado (o robado);
- iii. desde una fuente particular;



- iv. por un agente (estudiante o académico);
- v. sin la adecuada referencia;
- vi. con o sin intención de engañar.

Posteriormente, Pecorari (2008, pp. 98-110) elabora una definición de rasgos del plagio académico sobre la base de documentos elaborados por estudiantes de segundas lenguas que se encontraban en etapas adelantadas de su educación superior, al considerar que las dificultades demostradas se acercaban a las que evidencian los estudiantes nativos al ingresar al mismo nivel de enseñanza, en la medida en que se revelarían escasas estrategias para transformar el conocimiento que han leído anteriormente. Señala que para considerar una pieza textual como *plagio* se deben cumplir tres condiciones. En primer lugar, el texto creado debe contener palabras o ideas que también estén presentes en el texto de origen (que ha sido utilizado para la elaboración de la tarea). Luego, la similitud de las palabras o ideas no debe ser coincidente; aspecto que se determina, por lo general, a partir de la longitud de los conjuntos de palabras consecutivas que se repiten entre los documentos. Por último, el texto creado por el estudiante debe fallar en la atribución de la información en relación con los documentos que se emplean para la elaboración de la tarea. Por tanto, simplifica el modelo y lo define como: “The use of words and/or ideas from another source, without appropriate attribution” (p. 4).

Además, se reconoce, a partir de Moore Howard (1999, cit. en Pecorari, 2008), un tipo de plagio que no tendría la intención de engañar (propio de los aprendices de una segunda lengua y de los estudiantes que ingresan a una cultura disciplinaria) denominado *patchwriting*, definido como “copying from a source text and then deleting some words, altering grammatical structures, or plugging in one synonym from another” (cit. en Pecorari, 2008, p.5). Para los estudiantes que se encuentran en una etapa de ingreso al discurso propio de la disciplina (en otras palabras, los primeros años universitarios) el *patchwriting* es un paso necesario e inevitable, puesto que están aprendiendo a negociar su entrada a las nuevas comunidades discursivas, tal como precisa Sutherland-Smith (2008, p.

26): “Patchwriting is seen as a means of developing from apprentice to skilled writer, rather than as an issue of deliberate flouting of academic integrity guidelines”.

En la presente investigación, se trabajará con una definición de plagio académico que recoge aspectos tanto de Pecorari (2008) como de la caracterización de *patchwriting* de Howard (1999, cit. en Pecorari, 2008). Sin embargo, se espera que el resultado de los análisis permita dar luces sobre rasgos definitorios del plagio académico realizado por estudiantes que recién ingresan a alguna cultura disciplinaria, al relacionarlos, directamente, con la atribución de la información ajena implicados en las técnicas de intertextualidad revisadas: citas integrales y no integrales.

### 3. METODOLOGÍA

#### 3.1. TIPO DE ESTUDIO

La presente investigación es exploratoria, en la medida en que se constituye como una primera aproximación sobre la relación entre el *plagio académico* y la representación de la comprensión de múltiples documentos en estudiantes de primer año universitario de dos carreras de la Universidad de Chile. Por otro lado, incluye características propias de una investigación descriptiva, dado que intenta definir algunos rasgos de las formas de citación y de referencia a la información ajena entre alumnos de dos carreras con tradiciones disciplinarias diferentes: Licenciatura en Lengua y Literatura Hispánica (LLyLH) y Plan Común de Ingeniería (PCI).

#### 3.2. RECOPIACIÓN DE LA INFORMACIÓN

El total de textos analizados se consiguió a través de la elaboración de una tarea de escritura que debían realizar los estudiantes, como parte de las actividades evaluativas propias de una asignatura. El trabajo consistió en desarrollar una respuesta a un estímulo, con base en la lectura de dos textos facilitados por los profesores de los cursos. El tópico de estos documentos se centraba en aspectos propios de cada ramo (ver Anexo 2), con el fin de que la tarea fuera lo más natural posible, tanto en contenido como en forma. Con esto, se busca que los documentos resultantes entreguen datos provenientes de un contexto normal de evaluación.

Los instrumentos utilizados (ver Anexo 3) requieren la integración de la información proveniente de la lectura de los dos documentos de base. En este sentido, la correcta y completa consecución de la tarea implica, necesariamente, el uso explícito de ambas fuentes. Cada uno de ellos contó con la aprobación del coordinador general de la asignatura o con la del profesor de la cátedra<sup>23</sup>. Además, fueron validados tanto por el

juicio de un experto, como por un pilotaje de una pequeña muestra, cuyos resultados permitieron determinar su pertinencia para la presente investigación.

Por último, las respuestas de los estudiantes debían ser elaboradas en formato electrónico (en cualquier procesador de texto) y enviadas por el portal *u-cursos* o al correo electrónico del investigador.

### 3.3. SUJETOS DE ESTUDIO

Del universo total de alumnos que cumplían su primer semestre universitario durante el año 2010, se tomó en cuenta una muestra no probabilística de carácter accidental, a partir de dos carreras de la Universidad de Chile: Licenciatura en Lengua y Literatura Hispánicas (LLyLH) y Plan Común de Ingeniería (PCI). La elección de las carreras responde, principalmente, a la necesidad de investigar dos disciplinas con tradiciones discursivas completamente opuestas de una misma institución educativa. Por otra parte, la elección de la universidad se relaciona con, por un lado, que está catalogada como una de las universidades más prestigiosas del país y, por otro, por la facilidad de accesos a los profesores de cátedra por parte del investigador. Los sujetos estudiados fueron 104 estudiantes: 43 de la asignatura Lingüística estructural de LLyLH y 61 de la asignatura Introducción a la Ingeniería de PCI. Por último, la selección de las asignaturas se corresponde con las características que tienen las evaluaciones en cada una de ellas, esto es, evaluar contenidos en una tarea de escritura a partir de la lectura de documentos.

### 3.4. TRATAMIENTO DE LA INFORMACIÓN

Con el objetivo de procesar de manera efectiva y elaborar una metodología consistente para el almacenamiento de corpus de datos, se optó por almacenar cada documento en el formato XML, puesto que presenta una serie de ventajas en relación con otros formatos. En primer lugar, está basado en etiquetas (*tags*), que permiten jerarquizar los elementos que componen a los documentos. Además, se diferencia de otros lenguajes

masivos propios de la *Web*, como HTML, en la medida en que no tan solo *representa* la información, sino que también la estructura, en tanto separa contenido de formato. Por último, al estructurarse jerárquicamente, es posible realizar búsquedas solo de los elementos que se deseen estudiar, por medio de *Path Expression* (XPath, en el caso de los documentos XML), que, tal como indican Antoniu y Van Harmelen (2002, p. 45): “specifies how a node, or a set of nodes, in the tree representation of the XML document can be reached”.

Un documento XML está constituido por una serie de componentes detallados en el cuadro 13. Para ser procesado y compartido, cada documento XML debe estar *sintácticamente* bien construido, esto es, seguir una serie de reglas, como: presentar solo un prologo, inscribir los elementos entre una etiqueta de apertura y otra de cierre, procurar que los elementos contengan solo texto, entre otras.

Cuadro 13: *Componentes de un archivo XML*

<b>Carácter</b>	<b>Componente</b>	<b>Descripción</b>
<b>Obligatorio</b>	Prólogo	Declara su condición de XML y define el codificado (en nuestro caso, corresponde a UTF-8)
	Elementos	Son los componentes más importantes, es lo que <i>dice</i> un documento. Se identifican por una etiqueta de apertura y una etiqueta de cierre.
<b>No obligatorio</b>	Atributos	Especifican a los elementos.
	Comentarios	Explican ciertos pasos para la construcción del documento y no son procesados en los análisis.

*Nota 1.* Basado en Antoniu y Van Harmeleen (2002)

Si bien los nombres de las etiquetas de los elementos pueden ser creados libremente, se prefiere que éstas se unifiquen, con el fin de que pueda constituirse, de forma efectiva, como un formato uniforme de intercambio de datos. Para esto, las diferentes comunidades han establecido un vocabulario específico de dominio, como por ejemplo, las matemáticas (MathML), la bioinformática (BSML) o la astronomía (AML) (Antoniu y Van Harmeleen

(2002, p. 26). El etiquetado de textos digitales se realiza gracias al *Text Encoding Initiative* (TEI), que establece los parámetros esenciales para la construcción de documentos XML de corpus contruidos a partir de lenguaje natural.

En el presente trabajo, las etiquetas utilizadas se realizaron teniendo como base la última versión publicada por TEI (2010). En específico, se etiquetaron cada una de las citas realizadas por los estudiantes, diferenciando entre citas directas y citas indirectas. El criterio utilizado para diferenciarlas se centra en lo planteado por Maldonado (1999, p. 3554), quien apunta que la cita directa se identifica a partir de alguna marca tipográfica (comillas, guiones o cursiva), mientras que la cita indirecta se establece a partir de su subordinación con el verbo introductor, por medio de la partícula *que*. Para la visualización de un ejemplo de un documento XML realizado para esta investigación, véase el Anexo 4.

### 3.5. ANÁLISIS DE DATOS

El análisis de los datos se realizó en tres grandes etapas. La primera de ellas estuvo orientada a determinar, por medio del análisis computacional, las distancias semánticas latentes entre los textos creados por los estudiantes y los documentos entregados para el desarrollo de la tarea, puesto que a través de aquéllas es posible aproximarse al modelo situacional elaborado por cada alumno, a partir de la lectura de los textos base. Para lograrlo, antes de realizar el LSA, todos los documentos debieron ser pre-procesados, con el propósito de mejorar los resultados, por medio de la extracción de las *stopwords*, consideradas como las palabras más frecuentes (preposiciones, artículos, conjunciones, etc.) (Bécue-Bertaut, 2010, p. 23) y el *stemming*<sup>24</sup> (reducir las palabras a su raíz). El análisis LSA comienza con la eliminación de todas las palabras que ocurren una sola vez (*hapax logomema*), para reducir las dimensiones semánticas resultantes del análisis. Posteriormente, se construyó la matriz término/documento con el fin de establecer las distancias coseno derivadas del análisis realizado. Tal elaboración, se realizó considerando tipos como *n-grama 1*, vale decir, una sola palabra<sup>25</sup>. Estos valores se expresan entre 0 y 1, es decir, mientras más se aproxime dicha cifra a 1, mayor similitud semántica latente se

encontrará entre el texto del estudiante y el texto usado como estímulo, lo que deriva en una adecuada representación del modelo de situación del lector, producido gracias a la lectura de los documentos.

La segunda etapa se enfocó en determinar, de forma cuantitativa, la cantidad de citas presentes en los documentos, dado que, de acuerdo con Bazerman (2004) y Swales (1990), son los recursos más reconocibles (por tanto, objetivables) de intertextualidad. Este análisis se realizó considerando tanto la categorización proveniente de los estudios gramaticales, como la que deriva de los estudios sobre la escritura académica. En relación con esta última aproximación, también se consideraron otros aspectos relevantes que indican, efectivamente, atribución de la información ajena, pero que no son mecanismos de cita (por ejemplo, las notas al pie de página o la bibliografía).

La última etapa se centró en analizar los documentos en relación con los rasgos definitorios de presencia de plagio académico. Con este propósito, se utilizó, en primer lugar, las distancias entre los documentos y los textos base, para representar la cantidad de palabras similares que existe entre los textos creados y los originales. Posteriormente, para dar cuenta de los grupos de palabras similares, se realizó un segundo análisis computacional con LSA: esta vez, considerando los tipos como *n-grama* 2. Se probó con *n-gramas* superiores, sin embargo, los valores resultaron muy bajos como para que ayudaran en la decisión posterior. El empleo de *n-grama* superior a 1 descansa bajo el supuesto de que a mayor longitud de las unidades similares mayor probabilidad existe de encontrar plagio (Češka, 2009, pp. 9-10).

#### 4. ANÁLISIS Y PRESENTACIÓN DE LOS RESULTADOS

Con el fin de que este apartado sea lo más claro posible, cada uno de los resultados será presentado separando los datos obtenidos para los dos grupos estudiados, esto es, LLyLH y PCI.

##### 4.1. APROXIMACIÓN A LA REPRESENTACIÓN DE LA COMPRESIÓN DERIVADA DE LA LECTURA DE DOS DOCUMENTOS A PARTIR DE LSA

###### 4.1.1. VISUALIZACIÓN DE LA COMPRESIÓN DE DOCUMENTOS

Dado que las tareas planteadas implicaban la integración de la información proveniente de los dos textos base, se modelaron los datos con el fin de representarlos en un grafo que evidenciara la distancia entre los textos elaborados por los alumnos y los documentos leídos para la consecución de la tarea, tal como se observa en las Figuras 8 y 9.



Figura 8: *Distancia entre los documentos en LLyLH*

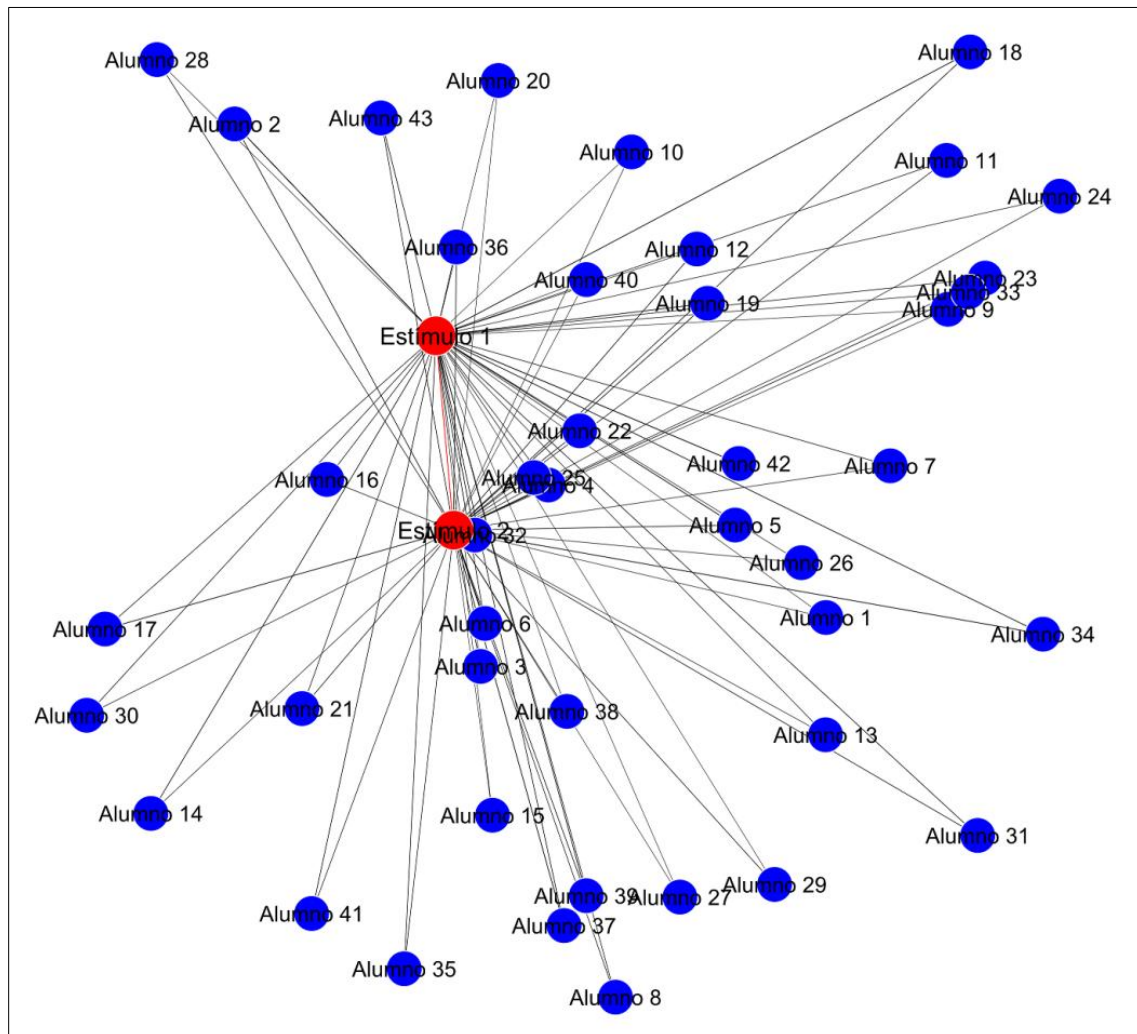
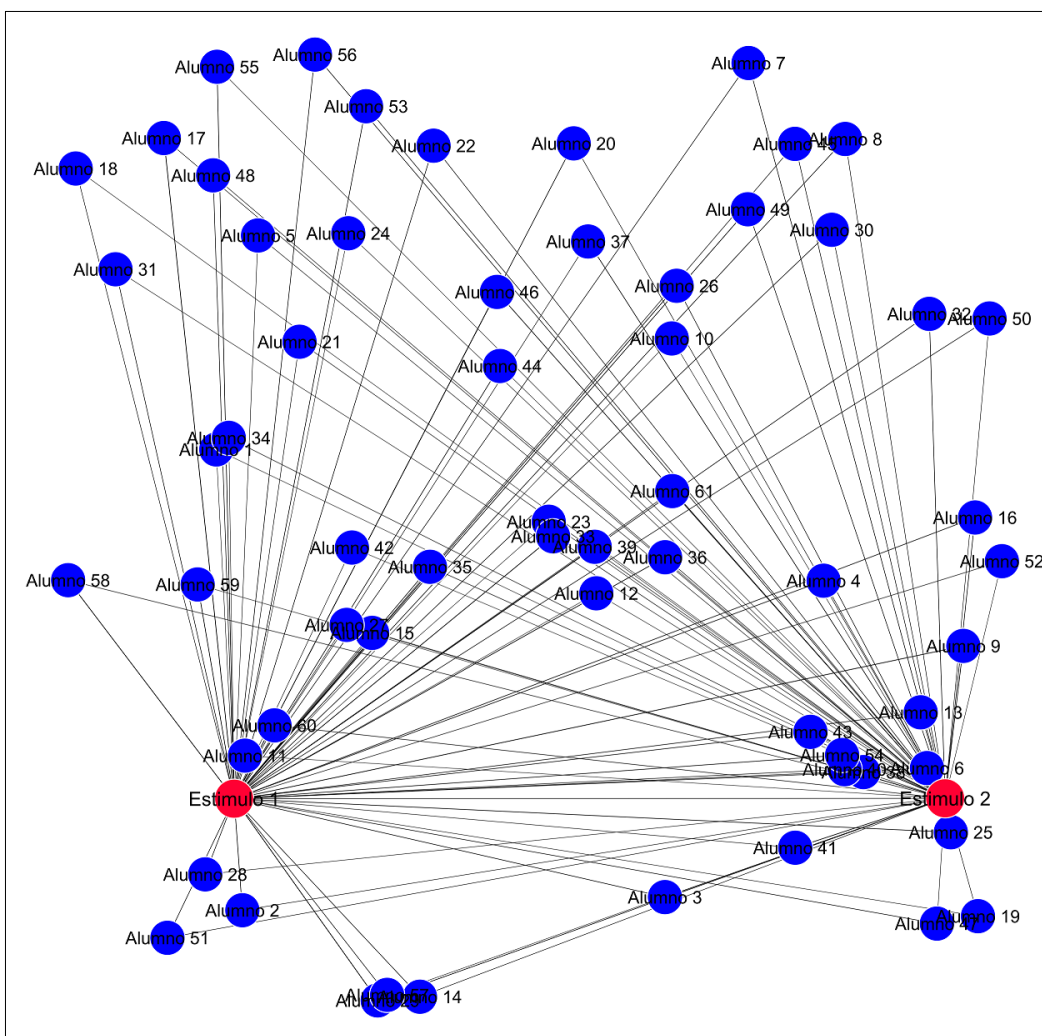


Figura 9: *Distancia entre los documentos en PCI*



En las Figuras 8 y 9, la red presentada se compone tanto de nodos, correspondientes a los documentos, como de vértices, que permiten la relación entre los nodos. Los nodos señalados con color rojo corresponden a los textos base, mientras que los azules se identifican con los trabajos de los estudiantes.

Aquellos nodos que se encuentran más alejados de los textos base presentan una menor relación semántica latente con éstos. La amplia dispersión visualizada en la Figura 9 demuestra que los documentos procesados de la carrera de PCI se adecuarían menos a los modelos situacionales derivados de la lectura de los textos base.

Por otro lado, la lejanía considerable de ciertos nodos se explicaría por el uso de fuentes ajenas a las solicitadas para desarrollar la actividad: sobre todo, *Wikipedia*<sup>26</sup>. En consecuencia, el análisis de distancia entrega luces sobre las prácticas de búsqueda y utilización de información confiable, dentro de la actividad universitaria (cf. Moya, L’Huillier, Velásquez, Ríos, 2010).

#### 4.1.2. INTEGRACIÓN DE LA INFORMACIÓN Y DISTANCIA TEMÁTICA ENTRE DISCIPLINAS

El tópicus discursivo de cada uno de los textos base leído por los estudiantes varía dependiendo de la disciplina en la que se inscriba. Dado que el tema de los documentos era similar según la asignatura, su similitud semántica latente indicará las cercanías entre los temas discursivos (cf. Kintsch, 2002). En cambio, la comparación realizada entre los documentos de las dos disciplinas indica que su cercanía es mucho menor (ver Cuadro 14). En cada caso, la comparación se realizó estableciendo *n-grama 1*, dado que solo se quiere determinar su grado de similitud y no posibles plagios derivados de la correlación de grupos de palabras consecutivas. Tal como se mencionó en 2.2.2., las similitudes se establecen por medio de la distancia *coseno*, expresada entre los valores 0 y 1, donde 1 representa la igualdad absoluta entre los documentos comparados.

Cuadro 14: *Comparación de las distancias LSA entre los textos de estímulo de cada asignatura*

	<b>Texto PCI 1</b>	<b>Texto PCI 2</b>	<b>Texto LLyLH 1</b>	<b>Texto LLyLH 2</b>
<b>Texto PCI 1</b>	1	0,55	0,24	0,20
<b>Texto PCI 2</b>	0,55	1	0,23	0,19
<b>Texto LLyLH 1</b>	0,24	0,23	1	0,66
<b>Texto LLyLH 2</b>	0,20	0,19	0,66	1

La representación de la comprensión de más de un documento se verifica a través de la integración de las fuentes de información implicadas en la elaboración de la tarea, por lo que resulta necesario establecer el nivel de uso de cada uno de los textos leídos en las tareas creadas por los alumnos. Una integración adecuada de información proveniente de ambas fuentes implica una construcción adecuada de la representación cognitiva de la tarea impuesta, esto es, se construiría un modelo situacional pertinente, a partir de la lectura de más de un documento. De esta manera, dicho nivel se puede caracterizar a partir de las distancias LSA definidas, en otras palabras, si el valor planteado es alto, el uso de aquella fuente de información es mayor (cf. Foltz, Britt y Perfetti, 1996). En las Figuras 10 y 11, se muestra la distribución de las distancias LSA de cada documento por el porcentaje de tareas realizadas por los estudiantes que se ubicaron en tales niveles.

Figura 10: *Gráfico de porcentajes de distribución de documentos según distancias coseno LSA, LLyLH*

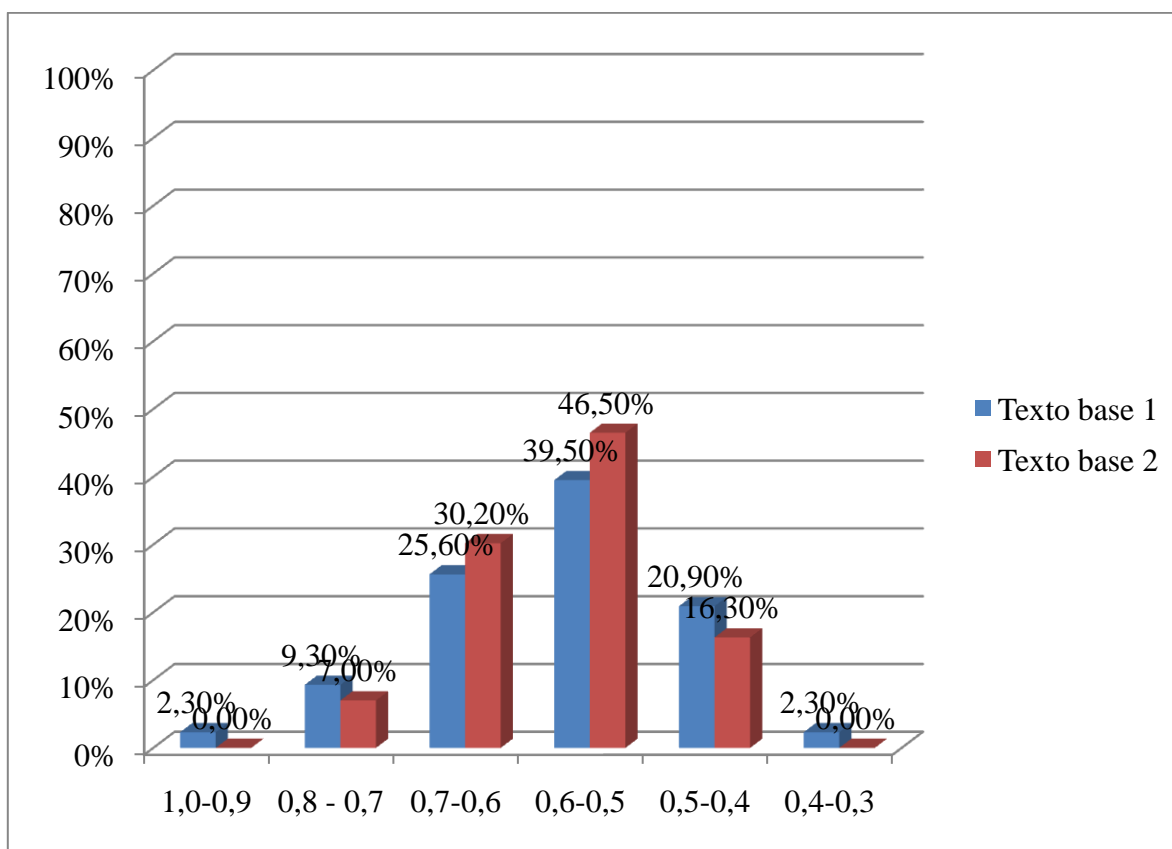
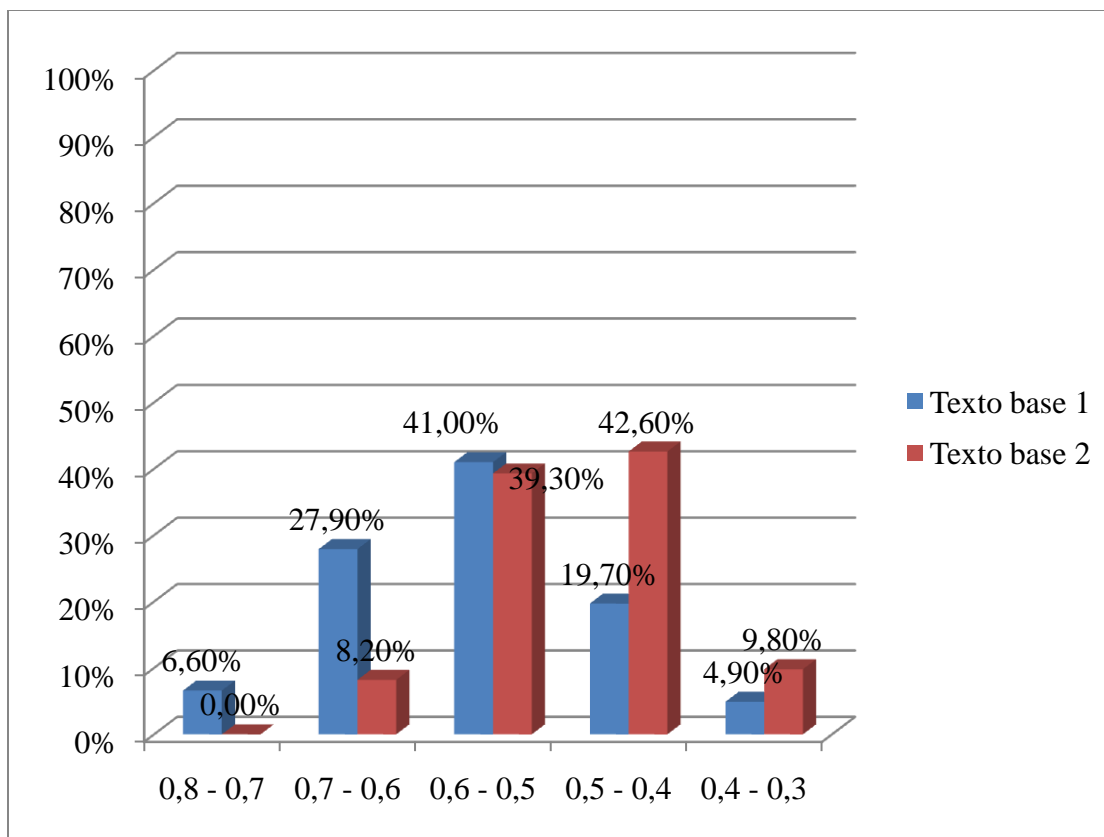


Figura 11: Gráfico de porcentajes de distribución de documentos según distancias coseno LSA, en PCI

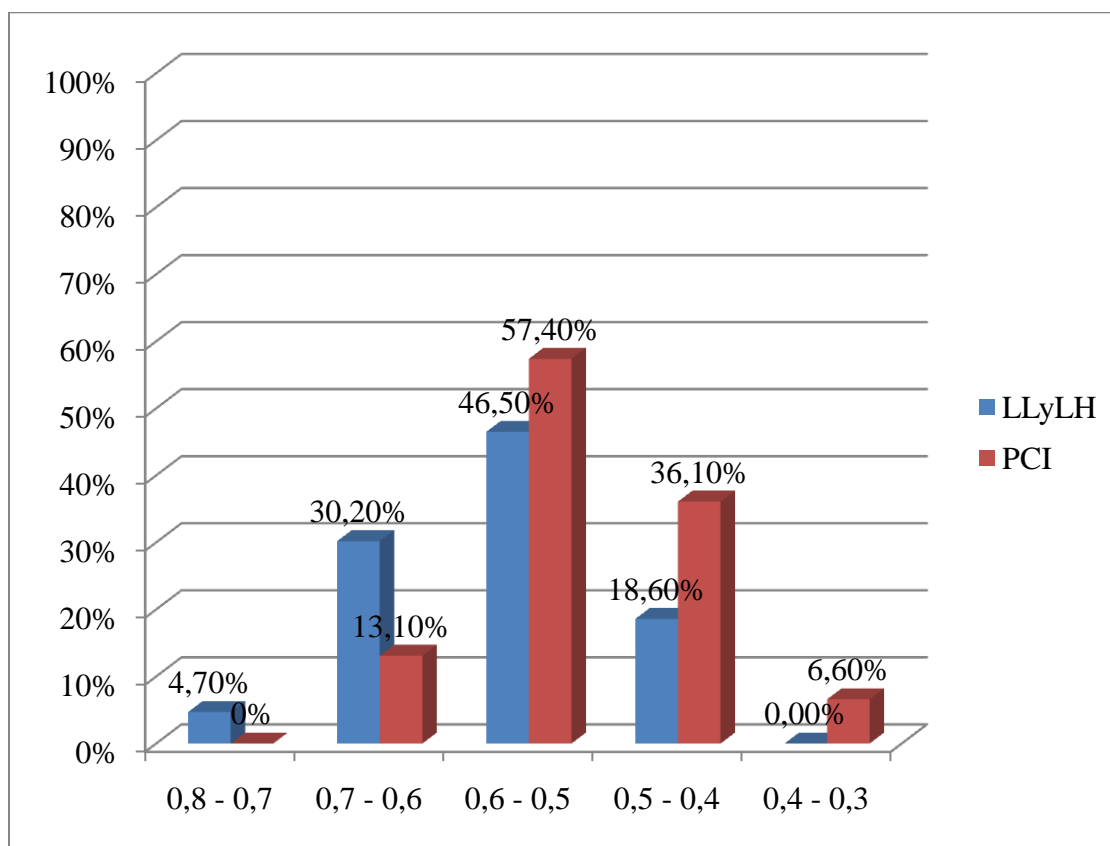


Como se puede apreciar, el uso de los documentos en LLYLH es bastante similar, debido a que no se evidencian grandes diferencias en la cantidad de alumnos que emplearon uno u otro texto para dar cuenta de la tarea. Además, es posible resaltar que los estudiantes, en su mayoría emplearon adecuadamente ambos textos, al situarse en las distancias que implican una alta similitud. En específico, el 76,7% de los documentos presentó una similitud sobre 0,5 con el texto 1, mientras el 83,7% de evidenció una similitud sobre 0,5 con el texto 2. En cambio, los resultados de uso de documentos para PCI señalan que existe una gran diferencia: utilizan preferentemente el texto 1, en la medida en que el 68,9% de los documentos presentaron una distancia entre 0,5 y 0,7 con él; mientras que el 52,4% de los documentos, presentó una baja similitud con el texto 2, entre 0,3 y 0,5, lo que ratifica su

escaso uso por parte de los estudiantes. En consecuencia, la preeminencia de uno sobre otro implicaría una consecución poco lograda de la tarea planteada.

Otra medida adecuada para aproximarse al nivel de comprensión de los textos a partir del análisis con LSA se relaciona el promedio de uso de los documentos entregados, tal como se particulariza en la Figura 12.

Figura 12: Gráfico de porcentajes de distribución de documentos según promedio de distancia coseno LSA entre los documentos



Como se observa, tanto en LLYLH como en PCI, la mayoría de los documentos creados por los estudiantes presentan una similitud alta (sobre 0,5) con el promedio de uso entre los dos textos base. Sin embargo, en el caso de PCI, es posible constatar un alto porcentaje de alumnos (42,7%) que presentan una similitud semántica latente baja (entre

0,3 y 0,5), considerando el promedio entre ambos textos, lo que se podría argumentar por el uso de fuentes de información diferentes a las entregadas para el desarrollo de la tarea (de la misma manera como se planteó en 4.1.1).

#### 4.2. INTEGRACIÓN DE LA INFORMACIÓN AJENA: CITAS

Tal como se mencionó anteriormente (véase Metodología, sección 3.5.), el recurso de integración de la información (intertextualidad) más sencillo de estudiar son las citas, en la medida en que es el más objetivable. Aun así, no todos los documentos presentan citas: el 25,5% de LLyLH y el 22,9% de PCI. Lo anterior no significa, *per se*, que no se utilice otra manera de atribución de la información, puesto que, en algunos casos, se evidencia empleo de bibliografía o notas al pie de página, pero no se constituyen como estrategias que permitan identificar claramente los segmentos que no pertenecen al propio autor. Concretamente, los textos de los alumnos de LLyLH suman 98 citas en total, mientras, por otro lado, los textos de PCI evidencian 126 citas. En el cuadro 15, se especifica la cantidad de palabras por cita que se encuentra en cada una de las disciplinas.

Cuadro 15: *Cantidad y porcentaje de palabras citadas según token frequency*

<b>Disciplina</b>	<b>Palabras en total</b>	<b>Palabras citadas</b>	<b>Porcentaje de palabras citadas del total</b>
<b>LLyLH</b>	29185	3789	13%
<b>PCI</b>	34157	2733	8%

#### 4.2.1. ANÁLISIS SINTÁCTICO

El primer análisis realizado de las citas encontradas buscó determinar qué tipo de citas predominaba en las dos disciplinas, tal como se detalla en el Cuadro 16.

Cuadro 16: *Porcentaje de uso de citas directas e indirectas por disciplina*

	<b>Citas directas</b>	<b>Citas indirectas</b>
<b>LLyLH</b>	83,7%	16,3%
<b>PCI</b>	63,5%	36,5%

Como se aprecia, en ambos casos se documenta un uso mayoritario de citas directas, pero su uso es ampliamente mayor en LLyLH en comparación con PCI. En específico, se correlacionan con lo que se postula en las investigaciones para el discurso académico en inglés: “Direct quotation, for example, is quite normal in the ‘soft’ fields” (Pecorari, 2008, p. 46). Para el caso de PCI es posible advertir que se presenta un porcentaje no menor de citas indirectas, lo que podría reflejar algún grado de inconsciencia en el empleo de estrategias de citación, en la medida en que la cita directa requiere más estrategias textuales y formales para conseguirse adecuadamente a los requerimientos formales de la disciplina. Por otro lado, dichas diferencias podrían originarse en la mayor preocupación por la información textual explícita que se manifestaría en LLyLH, por el tipo de formación que reciben. No obstante lo anterior, los estudiantes se encontraban en una etapa muy inicial de su formación (cerca del segundo mes de estadía en la universidad), por lo que no estarían lo suficientemente inmersos en su disciplina (ver ejemplos en Cuadro 17).



Cuadro 17: *Ejemplos de cita directa e indirecta en las disciplinas*

<b>Disciplina</b>	<b>Cita directa</b>	<b>Cita indirecta</b>
<b>LLyLH</b>	Bouknak dice: “Se puede afirmar que la técnica del trabajo de la piedra es un índice del desarrollo de las formas superiores del pensamiento (conceptos)”. (Sujeto 6)	En el texto de Bunak se da el ejemplo de los homínidos acheulenses que formaban palabras y los restos encontrados de esta época indican un cerebro mayor que los anteriores, una mandíbula más pequeña para poder articular mejor, entre otros. (Sujeto 25)
<b>PCI</b>	Como dice H. Gallegos: “Cuando se fabricó la bomba atómica o cuando el hombre llegó a la Luna se aplaudieron ambos logros como éxitos”. (Sujeto 15)	Lo que plantea el ingeniero Gallegos es que debe crearse una nueva era de la ingeniería donde exista un compromiso de utilizar de manera eficiente y consciente los recursos naturales, ya sea agua, madera, combustibles fósiles, etc. (Sujeto 14)

Asimismo, la cita directa como unidad de análisis debe estar introducida por algún mecanismo sintáctico/textual que permita, por un lado, identificar la procedencia o atribución de la información y, por otro, relacionar explícitamente dos segmentos textuales. Pese a lo anterior, se documentaron casos en que no existía introducción o atribución, ya sea como referencia a la cita o como alguna otra estrategia como nota al pie de página (ver Cuadros 18 y 19).

Cuadro 18: *Porcentaje de tipos de citas directas con y sin atribución*

	<b>Con atribución</b>	<b>Sin atribución</b>
<b>LLyLH</b>	92,7%	7,3%
<b>PCI</b>	93,8%	6,2%

Cuadro 19: *Ejemplos de cita directa sin estrategia sintáctico/textual introductora*

<b>LLyLH</b>	<b>PCI</b>
A medida que pasaban los años se fue adaptando a su medio y no sólo por sus rasgos físicos notamos esto, sino por el empleo de diferentes instrumentos y su intelecto. “La evolución del lenguaje no puede ser estudiada más que teniendo en cuenta las condiciones de vida de la sociedad primitiva y la necesidad que determinó la formación de palabras”. (Sujeto 3)	Lamentablemente el desarrollo ético no avanza al mismo paso, “junto con el crecimiento poblacional y el avance tecnológico, la ciudad ha acelerado la historia”. (Alumno 18).

Este tipo de discurso directo, es decir, sin atribución o frase introductora e identificable solo a partir de marcas tipográficas, corresponde a una de las variantes no canónicas del discurso directo<sup>27</sup> que se atestigua en la prensa diaria escrita (Brunetti, 2009, p. 32). No obstante, la atribución de dichos segmentos textuales, en la prensa diaria, se puede desprender gracias a una serie de elementos contextuales (noticias contiguas, imágenes, momento en el que ocurre la noticia, etc.) que cooperan con su interpretación posterior, por lo que no resulta completamente necesaria su presencia. Aun cuando en el discurso académico sea posible inferir la atribución o tan solo su presencia no resulte confundidora, no se estaría cumpliendo con las características básicas (en este caso formales) impuestas por cualquier disciplina, al momento de construir sus discursos. Esta intertextualidad propia de la prensa escrita presente en los trabajos de los sujetos indicaría

su bajo conocimiento sobre las características formales que articulan los géneros discursivos de su disciplina. Al carecer de estas habilidades, los estudiantes recurrirían a las características textuales de los géneros discursivos formales que efectivamente manejan, para lograr dar cuenta de la voz ajena en otros contextos formales a los que se están integrando.

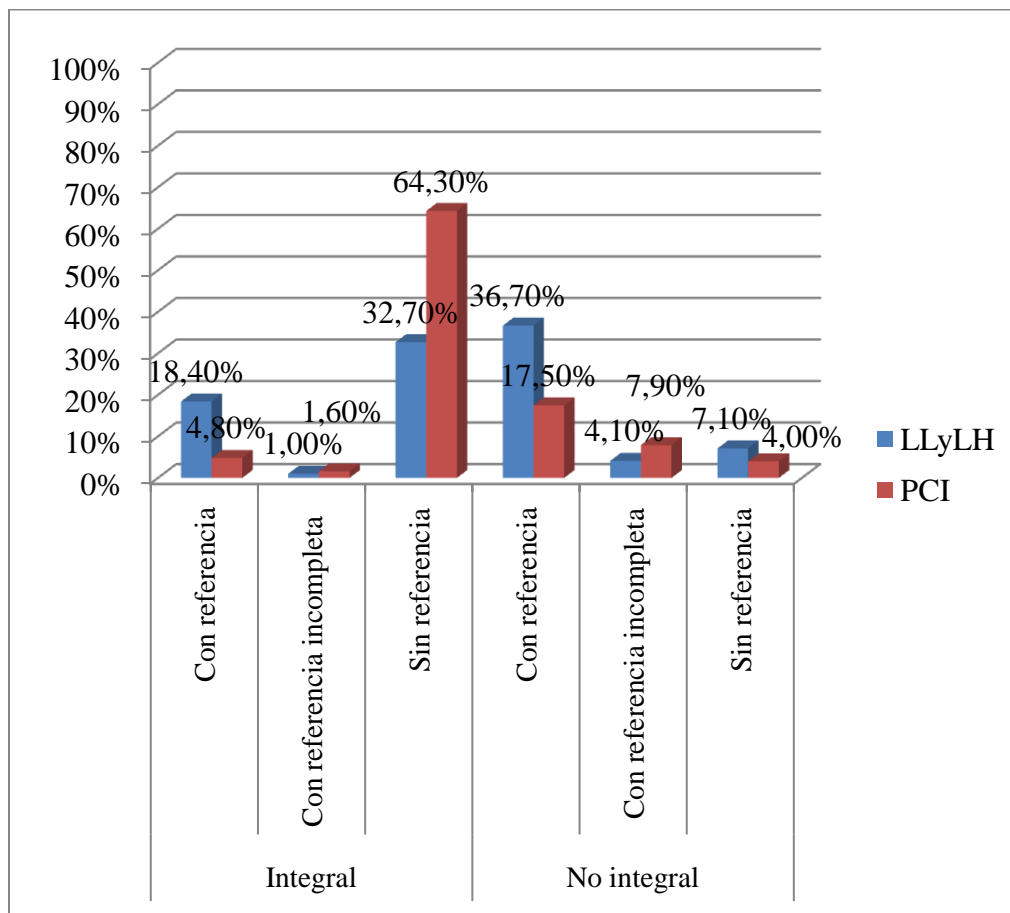
#### 4.2.2. ANÁLISIS DE CITAS A PARTIR DEL DISCURSO ACADÉMICO

La clasificación planteada por Swales (1990) entre citas integrales y no integrales fue establecida a partir de la investigación en artículos científicos incluidos en revistas arbitradas y con un editor en jefe, por lo que se soslaya el tema de la atribución de la información, en la medida en que no adquiere relevancia para sus pesquisas. De esta forma, dado que nuestro interés también radica en dar luces sobre la atribución de la información en estudiantes que se encuentran en un proceso de ingreso a la comunidad disciplinaria y a partir del análisis de los documentos, se estimó necesario ampliar dicha clasificación, considerando la manera en que ésta era atribuida por medio de algún mecanismo de referencia. Por tanto, en la presente investigación fue necesario establecer una gradación: con referencia, con referencia incompleta y sin referencia (ver Cuadro 20). Los resultados se detallan en la Figura 13. El detalle de cantidad de citas por sujetos se especifica en los Anexos 7 y 8.

Cuadro 20: *Caracterización de la gradación de las referencias propias de las citas*

<b>Tipo</b>	<b>Descripción</b>
<b>Referencia completa</b>	Debe explicitar, al menos, nombre de la fuente, el autor y la página desde donde se extrajo la información, en cualquier posición (luego de la cita, llamada al pie de página, final del documento).
<b>Referencia incompleta</b>	Debe explicitar el autor y el nombre de la fuente o la página desde donde se extrajo la información (luego de la cita, llamada al pie de página, final del documento)
<b>Sin referencia</b>	En el caso de las citas integrales, solo explicita el autor, dado que por definición debe indicarse el nombre del autor. En el caso de las citas no integrales, no existe ningún tipo de referencia.

Figura 13: Gráfico de porcentaje de tipos de citas según la clasificación planteada por Swales (1990) ampliada según el tipo de referencia implicada



Si bien ambos grupos presentan un mayor uso de citas integrales, PCI demuestra un mayor grado de citas sin referencias. Además, el porcentaje de citas con referencias incompletas es muy bajo, en las dos secciones. De esta manera, las citas integrales implican una noción de autoría, y por tanto, de reconocimiento a la voz ajena, en la medida en que es sintácticamente necesaria. Las citas no integrales, por su parte, requieren de un proceso consciente sobre los mecanismos de referencia de atribución de la información explícitos. Estos resultados no se relacionan con la caracterización llevada a cabo por Hyland (1999), quien asegura que las disciplinas científicas (*hard knowledge*) presentan un porcentaje dominante de citas no integrales (ver Cuadro 11). A pesar de que los textos utilizados en la

presente investigación son completamente diferentes a los analizados por Hyland (1999), el empleo mayoritario de citas integrales por parte de los estudiantes de PCI apoyaría la noción de que se encuentran en una etapa inicial de ingreso a la comunidad discursiva y de que ésta no imprimiría sus rasgos desde el comienzo de la formación académica del alumnado (cf. Soto, 2009) (ver ejemplos en el Cuadro 21).

Cuadro 21: *Ejemplos de tipos de citas integrales y no integrales por disciplina*

	<b>LLyLH</b>	<b>PCI</b>
<b>Cita integrada con referencia</b>	<p>Así detalladas por Bouknak: “Descenso de nivel en el punto lándico que caracteriza el aumento del lóbulo parietal del cerebro; redondeamiento del hueso frontal; disminución de la mandíbula, descenso de la línea milohioidea, que es índice del descenso de la laringe necesario para el desarrollo de la fonación oral, etcétera.” Referencia al pie de página: V.V. Bouknak. El origen del lenguaje. En los procesos de hominización, 1969. (Sujeto 1)</p>	<p>Thomas Tredgold define la ingeniería como “el arte de aprovechar los recursos de la naturaleza en beneficio del hombre y la sociedad”. Referencia al final del documento: Gallegos, Héctor (2004) La ética en la ingeniería. Sujeto 29).</p>
<b>Cita integrada con referencia incompleta</b>	<p>En su texto “Del grito a la palabra”, Bunak nos plantea, sobre el origen del lenguaje lo siguiente: “Las primeras palabras, que eran pocas y fundamentalmente monosilábicas, expresaban los hechos principales de la vida prehistórica”. (Sujeto 17).</p>	<p>De la página <a href="http://www.unbarrilmediolleno.com/archives/5188#more-5188">http://www.unbarrilmediolleno.com/archives/5188#more-5188</a> rescato el siguiente párrafo: ...”El estudio dice que Bea cree que sus hallazgos hasta el momento plantean “serias dudas acerca de las evaluaciones de riesgo utilizadas por BP y el Servicio de Gestión de Minerales, la agencia federal encargada de determinar si los planes de perforación eran adecuados”. (Sujeto 23).</p>

<b>Cita integrada sin referencia</b>	Tal como dice Bunak: “La diferencia principal entre el chimpancé y el hombre consiste en la imposibilidad de aquél para formarse nociones generales sobre un objeto”. (Sujeto 19).	Como el ingeniero del Perú, lo menciona es la única de las carreras que desempeña su trabajo entre los más fuertes, la naturaleza y sus acompañantes. (Sujeto 2).
<b>Cita no integrada con referencia</b>	Por ejemplo: “Al emplear nociones como ‘piedra’, ‘palo’, el hombre ya tiene una representación formada de numerosas propiedades de estos objetos y puede prever los resultados de su empleo”. Referencia al pie de página: (1) Bunak. V “Del Grito a la Palabra” (Sujeto 16).	En tercer lugar “la fábrica de productos y obras debe planearse en el sentido de se ejecuten sin causar el mínimo daño a la naturaleza”. Referencia al final del documento: La ética en la Ingeniería – Héctor Gallegos, pág. 12. (Sujeto 26).
<b>Cita no integrada con referencia incompleta</b>	Se establece el margen divisorio entre la condición de humano y animal se marca en el lenguaje, rasgo fundamental. Referencia al final del documento <a href="http://www.hiru.com/es/lengua_cas tellana/lengua_01400.html">http://www.hiru.com/es/lengua_cas tellana/lengua_01400.html</a> . (Sujeto 18).	Más que lo “técnicamente posible o económicamente rentable es humanamente deseable”. Referencia a continuación de la cita: (frase extraída del texto “Ética para Ingenieros”). (Sujeto 9).
<b>Cita no integrada sin referencia</b>	Así los sonidos “se convierten en símbolos de los conceptos relacionándose entre sí las percepciones de rasgos comunes de una categoría de objetos.” (Sujeto 18).	El concepto de ética, particularmente ‘la ética de la ingeniería’, la cuál es en conceptos generales “la moral reflexionada, sometida a la argumentación racional”. (Sujeto 48).



### 4.3. OTRAS CARACTERÍSTICAS

Junto con el análisis sobre las formas de citación, existen otros rasgos (relacionados por un lado con las citas y por otro con la atribución de la información) que nos permiten enriquecer la caracterización sobre algunos aspectos definitorios de la escritura de los alumnos al momento de ingresar a la educación superior, a saber: los introductores de cita integrales y el tratamiento de los referentes portadores de la voz ajena, la presencia explícita de bibliografía.

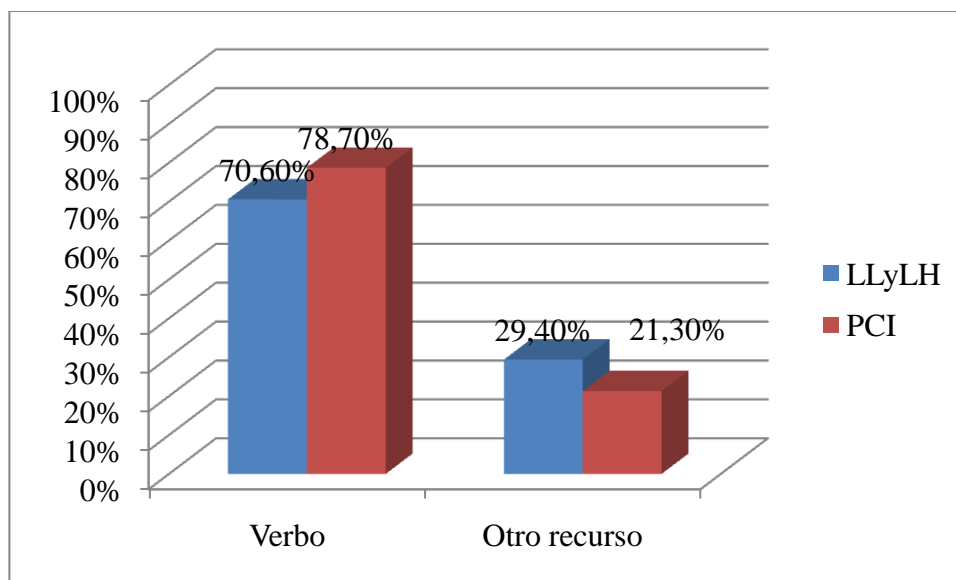
#### 4.3.1. INTRODUTOR DE CITA INTEGRAL

La literatura especializada precisa que las citas se integran en el discurso a partir de un *verbo dicendi* o algún otro que cumpla una función similar desde un punto de vista pragmático. No obstante, en nuestro corpus se detectaron otros recursos, aparte del uso de verbos (ver Figura 14). Tal como se detalla en los Cuadros 22 y 23, el verbo más utilizado para incorporar las citas integrales, en ambas disciplinas, corresponde a ‘decir’, esto es, el *verbo dicendi* semánticamente más amplio, por lo que adquiere significados trasladados de otros verbos que cumplirían funciones similares. Por otro lado, en los dos grupos, ‘según’ es el recurso más utilizado para referir a la voz ajena, sin ser verbo. Para ambos casos (verbos y otros recursos) se especifican algunos ejemplos en los Cuadros 24, 25 y 26.

Este tipo de recursos indicaría la actitud epistémica de los autores frente al conocimiento, que se refleja en la manera en que los escritores expresan sus opiniones frente al conocimiento científico. Este carácter evaluativo del discurso académico de los estudiantes (que implica una interacción entre los miembros de la comunidad) se manifiesta a través tanto de las nociones *postura* como de *compromiso*. La *postura* se refiere a una dimensión actitudinal que incluye características referidas a las formas en que los escritores presentan sus juicios u opiniones, mientras que el *compromiso* se relaciona con una dimensión de alineación, donde los escritores reconocen y se conectan con otros miembros de su comunidad (Hyland, 2005). Dentro de la dimensión actitudinal, se precisan dos

elementos esenciales para expresar los juicios de los autores: los *hedges* y los *boosters*. Mientras los primeros se refieren, prototípicamente, a los mecanismos que le permiten al escritor modificar el valor de verdad de sus proposiciones (Soto, 2002, siguiendo a Markkanen y Schöereder, 1997), los *boosters* (intensificadores), por el contrario, permiten al autor expresar seguridad en lo que están señalando (Hyland, 2005, p. 179). De esta manera, se reafirma la noción de que los estudiantes de primer año se encuentran en una etapa inicial de acceso a las diferentes comunidades discursivas, en la medida en que se emplean, mayoritariamente, recursos epistémicamente neutros (como *decir*, *plantear* o *según*), empleados para transmitir conocimiento establecido o verdades dentro de las comunidades, tal como asegura Soto (2002): “Mientras los conocimientos establecidos se comunican como verdades, las nuevas propuestas presentan frecuentemente *hedges* que modalizan epistémicamente las proposiciones, mitigando el compromiso de los autores con el valor de verdad de las mismas”.

Figura 14: Gráfico de porcentaje de tipo de introductor de cita integral



Cuadro 22: *Verbos introductores de citas integrales en LLyLH*

<b>Verbo</b>	<b>Cantidad</b>	<b>%</b>
decir	11	30,6
plantear	3	8,3
explicar	3	8,3
citar	3	8,3
afirmar	2	5,6
expresar	2	5,6
hacer+FN	2	5,6
referir	2	5,6
detallar	1	2,8
asegurar	1	2,8
llamar	1	2,8
dar	1	2,8
proponer	1	2,8
sintetizar	1	2,8
dar cuenta	1	2,8
señalar	1	2,8

Cuadro 23: *Verbos introductores de citas integrales en PCI*

<b>Verbo</b>	<b>Cantidad</b>	<b>%</b>
decir	14	20,0
plantear	7	10,0
mencionar	6	8,6
señalar	6	8,6
definir	5	7,1
citar	3	4,3
hablar	3	4,3
denominar	2	2,9
referir	2	2,9
proponer	2	2,9
basar	1	1,4
tomar+fn	1	1,4
afirmar	1	1,4
agregar	1	1,4
establecer	1	1,4
tomar	1	1,4
acotar	1	1,4
hacer+FN	1	1,4
constar	1	1,4
esbozar	1	1,4
parfrasear	1	1,4
proclamar	1	1,4
preguntar	1	1,4
aludir	1	1,4
rescatar	1	1,4
considerar	1	1,4
resumir	1	1,4

recordar	1	1,4
sugerir	1	1,4
plasmar	1	1,4

Cuadro 24: *Otros recursos para la integración de la voz ajena en LLyLH*

<b>Recurso</b>	<b>Cantidad</b>	<b>%</b>
Según	8	61,5
Para	3	23,1
En palabras de	1	7,7
Construcción sintáctica integradora	1	7,7

Cuadro 25: *Otros recursos para la integración de la voz ajena en PCI*

<b>Recurso</b>	<b>Cantidad</b>	<b>%</b>
Según	13	72,2
De acuerdo con	1	5,6
En otras palabras	1	5,6
Construcción sintáctica integradora	3	16,7

Cuadro 26: Ejemplos de recursos empleados para integrar la voz ajena

	<b>LLyLH</b>	<b>PCI</b>
<b>Recursos verbales</b>	Como dice Bouknak en la página 96 del texto “El origen del lenguaje”: “La idea general, cuyos elementos principales son los conceptos, es decir, representaciones de cosas y de fenómenos unidas por ciertos índices comunes”. (Sujeto 17).	Como dice Héctor Gallegos, “la técnica social por excelencia” (Sujeto 59).
	Como lo plantea Bunak, “la percepción de la acción y la del objeto o sujeto de la misma” (Sujeto 32).	El texto plantea que la primera vista puede que no se percaten de los problemas que puede traer la creación de algún elemento útil en al hombre. (Sujeto 42).
<b>Otros recursos</b>	Según Victor Bunak, “es capaz de elaborar imágenes mentales coherentes de objetos y acciones distintas, de diferenciarlas entre sí y de combinarlas con otras” (Sujeto 32).	Según Héctor Gallegos el “asegurar que dichos recursos sean destinados al bien común y accesibles a las generaciones futuras”. (Sujeto 45).
	Para V. Bouknak esto no es necesariamente así: “Según todas las probabilidades, a las particularidades del tipo craneológico de los diferentes grupos de los paleonthropus correspondían diferencias de pensamiento y de la palabra, pero estas diferencias no tienen inmediatamente su reflejo neto en la técnica del trabajo de la piedra”. (Sujeto 22).	De acuerdo con la definición de Tredgold, la ingeniería es “el arte de aprovechar los recursos de la naturaleza en beneficio del hombre y la sociedad”. (Sujeto 17).

#### 4.3.2. TRATAMIENTO DE LOS REFERENTES

La última caracterización sobre algunos rasgos textuales se relaciona con el tratamiento de los referentes portadores de la voz ajena. Como todo proceso de adecuación léxica, en la escritura académica se espera que la repetición se evite por medio de diferentes mecanismos. A pesar de lo anterior, en el caso del tratamiento de los referentes propios de la atribución de la información, se deben seguir ciertas normas que demuestren, por un lado, la comprensión de lo que se está señalando y, por otro, el uso de ciertas normas implícitas impuesta por las disciplinas. En este sentido, tal como se especifica en los Cuadros 27 y 28, y ejemplifica en el Cuadro 29, la amplia variedad en la selección léxica empleada demuestra la vacilación e indeterminación por las formas canónicas disciplinarias para dar cuenta de los referentes adecuadamente.

Cuadro 27: *Recursos para el tratamiento de referentes en LLyLH*

<b>Recurso</b>	<b>Cantidad</b>	<b>%</b>
Solo apellido	22	71,0
Nombre + apellido	2	6,5
Texto de + apellido	3	9,7
Autor	1	3,2
Inicial del nombre + apellido	2	6,5
Dicho	1	3,2

Cuadro 28: Recursos para el tratamiento de referentes en PCI

<b>Recurso</b>	<b>Cantidad</b>	<b>%</b>
Solo apellido	28	31,8
Nombre + apellido	21	23,9
Sinónimo	6	6,8
Nombre del libro	8	9,0
Texto	4	4,5
Pronombre personal	3	3,4
Nombre propio corporativo	3	3,4
Nombre	2	2,3
Palabras de + apellido	2	2,3
Señor + apellido	1	1,1
Frase de + autor	1	1,1
Inicial + apellido	1	1,1
Texto + nombre libro	1	1,1
Párrafo	1	1,1
Libro	1	1,1
Extracto	1	1,1
Fuente	1	1,1
Autor del texto	1	1,1
Fragmento + nombre libro	1	1,1
Frase	1	1,1



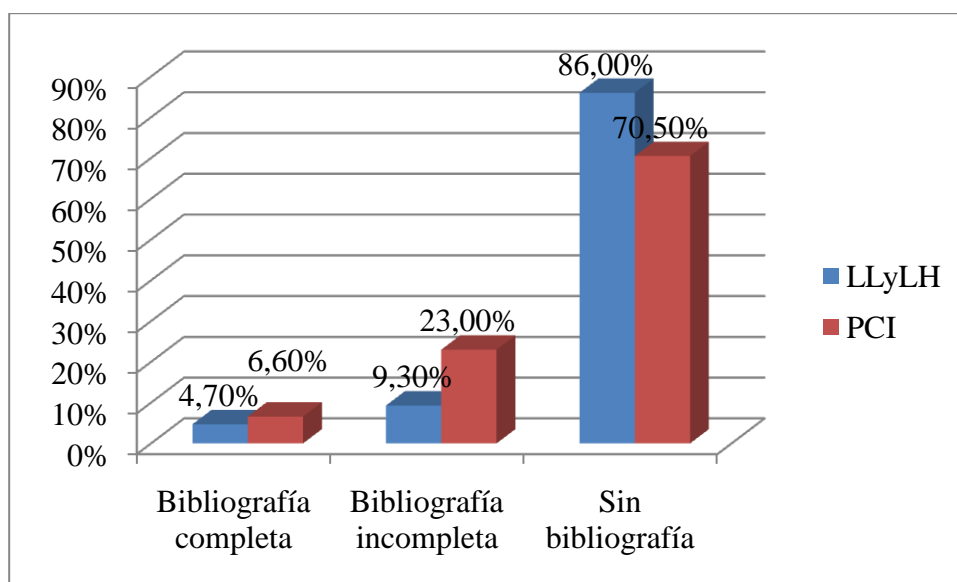
Cuadro 29: *Ejemplos de recursos para el tratamiento de referentes*

LLyLH	PCI
<p>V. Bouknak propone que “la relación entre la evolución del lenguaje y la evolución del intelecto, del empleo de instrumentos, del tipo físico, abre un vía metódica para la solución del problema del origen del lenguaje” (Sujeto 31).</p>	<p>Lo que plantea el ingeniero Gallegos es que debe crearse una nueva era de la ingeniería donde exista un compromiso de utilizar de manera eficiente y consciente los recursos naturales, ya sea agua, madera, combustibles fósiles, etc. y ceñirse a la necesidad del ser humano al momento de diseñar nuevas herramientas o nuevos métodos para ocuparlos en la vida diaria, eso es lo que el definió como “Eco-ingeniería”, un compromiso ético para con nosotros, las próximas generaciones y la naturaleza. (Sujeto 14).</p>
<p>Según explica el autor, la capacidad de un individuo para generar una imagen mental de objetos o acciones que son distintos entre sí, diferenciarlos de acuerdo a sus características generales y combinarlos con otras. Básicamente se refiere a que el hombre ya tiene la capacidad de distinguir y relacionar, aunque sea a grandes rasgos, objetos distintos entre sí. (Sujeto 17).</p>	<p>Como muy bien dice Héctor destruir la naturaleza es destruir a la humanidad. (Sujeto 17).</p>

### 4.3.3. PRESENCIA EXPLÍCITA DE UN APARTADO DE BIBLIOGRAFÍA

Dentro de los informes universitarios es común que se incluya una sección relativa a la bibliografía empleada durante la elaboración del trabajo, la que, muchas veces, se constituye como el único medio que atribuye algún tipo de información, que, no obstante, no cumple a cabalidad el objetivo de asignar el referente a determinada información. De esta manera, su propósito es citar “trabajos que sirvieron de fundamento o son útiles para una lectura posterior” (APA, 2001, p. 223). Los resultados de esta revisión se visualizan en la Figura 15.

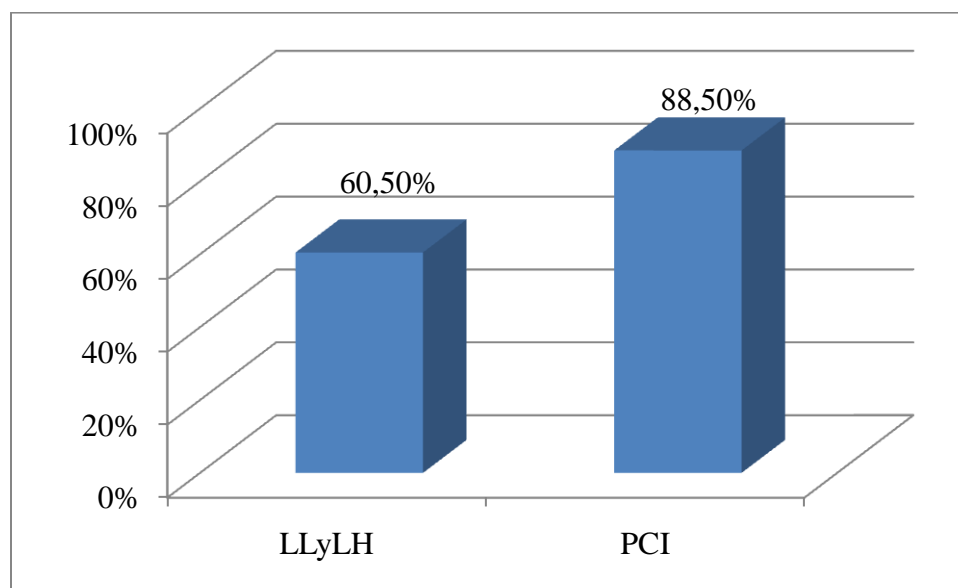
Figura 15: *Gráfico de porcentaje de documentos con bibliografía explícita en ambas disciplinas*



#### 4.4. PRESENCIA DE PLAGIO ACADÉMICO

Tal como se especificó en 2.4.3., la identificación del *plagio académico* requiere el cumplimiento de tres requisitos fundamentales: palabras o ideas similares al texto base, grupos de palabras o ideas similares y fallar en la atribución de la información (Pecorari, 2008). Para este análisis, se utilizaron las distancias LSA con *n-grama 1* (ver anexo 5 y 6) – para verificar la presencia de palabras o ideas similares– y con *n-grama 2* (para verificar la presencia de grupos de palabras o ideas similares). En este último caso, las distancias (ver Anexo 5 y 6) podrían parecer muy bajas en relación con *n-grama 1*, sin embargo, se debe recordar que el término utilizado para construir la matriz considera a dos palabras en lugar de una. En consecuencia, a pesar de que los valores sean bajos, indican, de igual forma, grupos de palabras que coexisten en ambos documentos. Por otro lado, en torno a la mala utilización de estrategias de referencia de la información, bastó con que una sola cita presentase falencias para considerarla un error. Los resultados sobre la presencia de plagio académico (siguiendo la definición propuesta por Pecorari, 2008) en los textos revisados se precisa en la Figura 16.

Figura 16: Gráfico de porcentaje de textos plagiados en LLyLH y PCI siguiendo a Pecorari (2008)



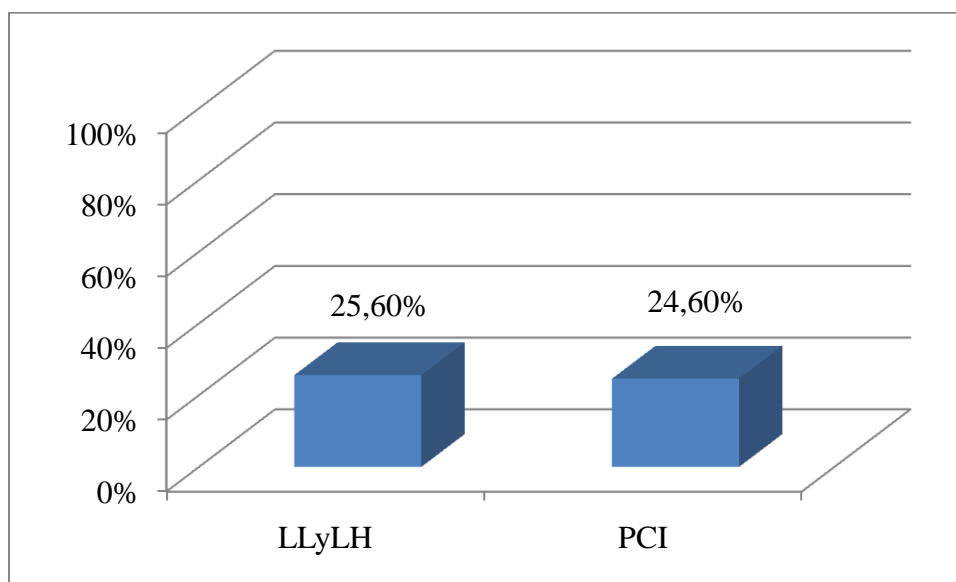
No obstante, el análisis realizado a partir de la definición entregada por Pecorari (2008) no considera el tiempo de permanencia en la enseñanza superior como un rasgo relevante al momento de determinar si un texto está *plagiado*. Los estudiantes en el primer año universitario se encuentran en un etapa en la que las herramientas para referirse a la información proveniente de otras fuentes aún no han sido entregadas (cf. Carlino, 2005). Además, como se revisó en 4.2.2., la categoría de citas establecidas por Swales (1990) no resulta suficiente para dar cuenta de las formas de transmisión del conocimiento ajeno empleadas por los estudiantes de primer año. En consecuencia, así como se considera el *patchwriting* como un paso ineludible dentro del proceso de adquisición de la escritura dentro del nivel académico, los mecanismos formales de intertextualidad propios de la disciplina también deben ser comprendidas como un *proceso de adquisición de formalidades disciplinarias*. Para efectos de nuestra investigación, no se puede juzgar de plagio a quien haya tenido la intención de atribuir información de manera identificable, en al menos un caso durante el desarrollo de su trabajo. Dado que las citas integrales siempre manifiestan un grado de atribución de la información (por definición), estimamos que la sola presencia de una de ellas ayudaría a determinar la *conciencia* que tiene el estudiante sobre la información que no es propia, cuyo origen debe ser establecido de alguna forma.

En concreto, a partir del análisis, planteamos que los rasgos determinantes del plagio académico para los alumnos que cumplen su primer año universitario serían los siguientes:

- Presencia de palabras o ideas idénticas entre la fuente de información y el texto elaborado;
- Presencia de grupos de palabras o ideas idénticas entre la fuente de información y el texto elaborado; y,
- Ausencia de citas; o,
- Presencia de citas no integrales sin referencias.

Por otro lado, la sola presencia de una sección de bibliografía no bastaría para identificar claramente la procedencia específica de ciertos bloques informativos (de la misma manera en que se manifiesta en 4.3.3.). En este sentido, los resultados provistos a partir del análisis (ver Figura 17) permiten identificar que, mayoritariamente, existe conciencia sobre la atribución de la información. Cabe señalar que la atribución establecida cubrió un amplio espectro de formas, como se detalló en 4.3.3.

Figura 17: *Gráfico de presencia de plagio, de acuerdo con criterios definitorios para alumnos de primer año*



#### 4.5. RELACIÓN ENTRE PLAGIO ACADÉMICO Y REPRESENTACIÓN DE LA COMPRESIÓN DEL DISCURSO

Dentro de la literatura especializada revisada, no se encuentran referencias que se aboquen a la relación que se podría establecer entre la representación de la comprensión de lectura (esto es, la elaboración de un modelo situacional) y la presencia de plagio académico. En el presente trabajo, los resultados de esta correlación (ver Cuadro 30 y Cuadro 31) indican que el promedio de similitud semántica latente de los textos plagiados con los documentos utilizados para responder al estímulo son más bajos en PCI que en LLyLH (ver Figura 18), lo que se identificaría con representaciones de la comprensión lectora menos adecuadas a los requerimientos de la tarea.

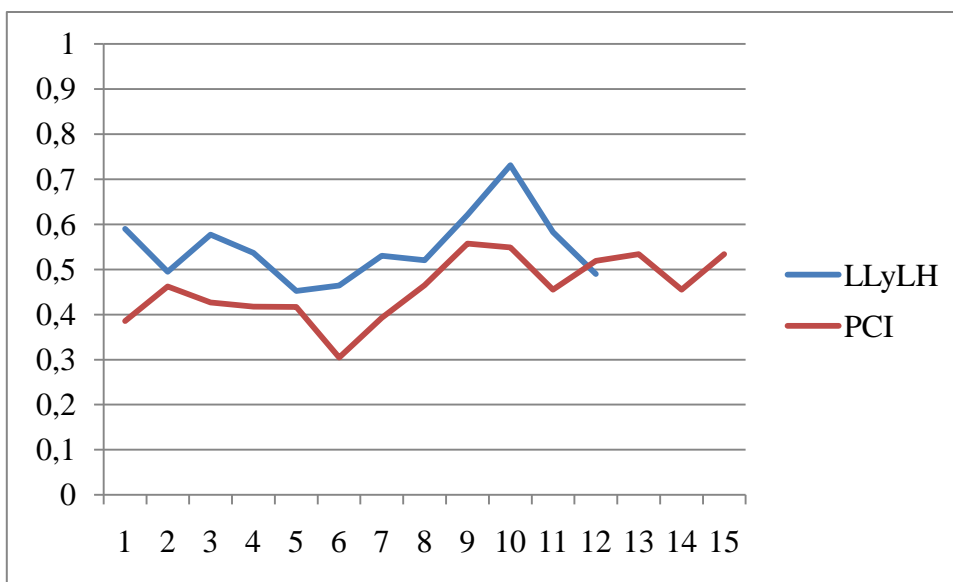
Cuadro 30: *Promedios de uso LSA en textos plagiados en LLyLH*

<b>Sujeto</b>	<b>Promedio coseno de uso de documentos</b>
Alumno 7	0,59012798
Alumno 9	0,49450453
Alumno 12	0,5773202
Alumno 13	0,53662753
Alumno 15	0,45198454
Alumno 27	0,46415448
Alumno 28	0,53029769
Alumno 29	0,52004581
Alumno 33	0,62134647
Alumno 34	0,73104896
Alumno 35	0,58285658
Alumno 40	0,49003646
<b>Promedio</b>	<b>0,54919594</b>

Cuadro 31: *Promedios de uso LSA en textos plagiados en PCI*

<b>Alumno</b>	<b>Promedio coseno de uso de documentos</b>
Alumno 1	0,38549982
Alumno 5	0,46202206
Alumno 6	0,42656949
Alumno 10	0,41720738
Alumno 11	0,41652327
Alumno 12	0,30463781
Alumno 24	0,39248126
Alumno 27	0,46532334
Alumno 37	0,55751106
Alumno 38	0,54851088
Alumno 39	0,45534918
Alumno 50	0,51895735
Alumno 51	0,53355445
Alumno 52	0,45507899
Alumno 61	0,53393071
<b>Promedio</b>	<b>0,45821047</b>

Figura 18: Gráfico de comparación de promedio de distancias LSA en textos plagiados en LLyLH y PCI



Por otra parte, es importante comparar los promedios alcanzados por los textos plagiados en el contexto global de cada carrera. Así, como se muestra en el Cuadro 32 y en el Cuadro 33 más de la mitad de los casos de plagio en cada disciplina se encuentran en el tercio que agrupa los documentos que tienen mayores distancias LSA con los textos leídos para responder el estímulo, es decir, con los documentos que no logran un modelo de situación adecuado, derivado de la lectura de múltiples documentos.

Cuadro 32: Porcentaje de distribución de textos plagiados según distancia LSA en LLyLH

División	Porcentaje de textos plagiados
0,6 – 0,79	18,2%
0,54 – 0,6	27,3%
0,45 – 0,54	54,5%



Cuadro 33: *Porcentaje de distribución de plagio según distancia LSA en PCI*

<b>División</b>	<b>Porcentaje de textos plagiados</b>
0,57 – 0,68	0,0%
0,5 – 0,56	33,3%
0,3 – 0,5	66,7%

De la misma forma en la que se puede establecer una relación entre integración de los documentos (que refleja el modelo situacional a partir de ambas lecturas) y el plagio, es posible determinar una correspondencia entre el uso de citas y la representación de la comprensión del discurso alcanzada (ver Cuadro 34 y Cuadro 35), en otras palabras, mientras más adecuada sea la representación cognitiva derivada de la lectura de fuentes múltiples mayor será la cantidad de citas presente en los textos. En ambas carreras, el porcentaje mayoritario de citas se encuentra en el tercio que agrupa los documentos que tienen distancias LSA menores que el resto de su conjunto.

Cuadro 34: *Porcentaje de distribución de citas según distancia LSA en LLyLH*

<b>División</b>	<b>Porcentaje de citas</b>
0,6 – 0,79	50%
0,54 – 0,6	32,7%
0,45 – 0,54	17,3%

Cuadro 35: *Porcentaje de distribución de citas según distancia LSA en PCI*

<b>División</b>	<b>Porcentaje de citas</b>
0,57 – 0,68	50,8%
0,5 – 0,56	26,2%
0,3 – 0,5	23%

## 5. CONCLUSIONES

En la investigación exhibida, de tipo exploratorio, se intentó entregar luces sobre la manera en que se relaciona, en alumnos que se encuentran cursando el primer año de universidad, la integración de proveniente de dos fuentes informativas en un modelo situacional adecuado y coherente tanto con la tarea propuesta como con las estrategias destinadas a la transmisión del conocimiento ajeno, en especial, las citas. Para lograrlo, por un lado, se presentó una aproximación a los modelos situaciones a partir del análisis con LSA; por otro, se describieron los procesos de cita, como facilitadores y portadores de la atribución informativa. De esta manera, se puede concluir:

1. Mientras los alumnos de LLYLH tendieron a presentar un mayor nivel de integración a partir de las fuentes, los estudiantes de PCI evidenciaron el uso prominente de una sola fuente, lo que implicaría la construcción dispar de modelos situacionales en relación con el campo disciplinario. Por otro lado, el análisis LSA, que demuestra una aproximación a la representación cognitiva construida, implica, por definición, el modelamiento del *conocimiento* que tiene el alumno sobre el tema escrito. Por tanto, un texto definido por las características de los instrumentos utilizados en nuestra investigación indicaría el nivel de *aprendizaje* de los alumnos (Foltz, Britt y Perfetti, 1996), puesto que, tal como mencionan Kintsch (1998) y Kintsch y Kintsch (2005), el almacenamiento de los modelos situacionales se ubica en la memoria de largo plazo. De esta manera, se garantiza que la comprensión sobre los discursos leídos se mantenga durante el tiempo. Sin embargo, la adecuada construcción de un modelo situacional que facilite un buen aprendizaje no depende tan solo del estudiante, sino que está en manos, principalmente, del profesor quien entrega la tarea que se debe desarrollar. Dichas tareas debiesen promover, de forma explícita, la integración de la información desde los documentos, gracias al empleo de preguntas con contenido argumentativo (Wiley y Voss, 1999) o aquellas que contengan investigación, revisión crítica o reflexión (Bhattacharya y Jorgensen, 2008, p. 203). Por último, la escritura no es solo la simple transmisión de ideas, sino que, adherimos lo propuesto por Carlino (2003, p. 411): “La escritura alberga un potencial epistémico, es

decir, no resulta sólo un medio de registro o comunicación sino que puede devenir un instrumento para desarrollar, revisar y transformar el propio saber”.

2. En cuanto al uso de citas, los resultados indicaron que los alumnos de LLyLH emplean más citas directas que los alumnos de PCI. Lo anterior concuerda con las características de las citas en las diversas disciplinas (Pecorari, 2008; Hyland, 2000, pp. 29-30). Por otro lado, si bien la dicotomía *integral/no integral* planteada por Swales (1990) es útil para nuestra investigación, no se hace cargo de casos que se pueden encontrar en contextos académicos que no correspondan a las publicaciones con comité editorial. Por tanto, proponemos considerar las referencias en cada una de las categorías señaladas, puesto que por una parte, creemos que son fundamentales para escribir posteriores problemas relacionados con la atribución adecuada de la información ajena; y, por otra, entendemos cualquier aspecto de la escritura académica como un proceso de aprendizaje que se visualiza como un *continuum* y no como un producto que se adquiere de una vez y para siempre (Carlino, 2005). En este sentido, los alumnos de PCI presentaron un mayor porcentaje de citas cuya referencia no era explicitada (68,3% contra un 39,8% de LLyLH). Por otro lado, el porcentaje de citas integrales para PCI (70,7%) no se relacionó con lo planteado por la literatura especializada (Hyland, 2000, p. 30), lo que se podría interpretar de dos maneras diferentes. Primero, los estudiantes, efectivamente, se encuentran en una etapa inicial en el ingreso a la comunidad discursiva, por lo que desconocen las normas implícitas derivadas de lo que se considera escritura adecuada en ese contexto. Segundo, las caracterizaciones realizadas sobre los rasgos escriturales (en nuestro caso, relativos a las formas de citación) se han desarrollado, principalmente, en inglés, por lo que las características entre las lenguas podrían variar. En cualquier caso, apremian descripciones claras sobre las maneras de integración de la información en español para todas las disciplinas, con el fin de establecer patrones específicos que permitan acercarse a una alfabetización académica eficaz, desde el primer año universitario. De esta manera, aunque concordamos con la postura de la escritura a través del currículum (Carlino, 2003), consideramos, también, esencial contar con una caracterización acabada sobre los rasgos que lo determinan, derivada de estudios lingüísticos sobre una gran base documental.

3. Las características revisadas relativas a la presencia de un apartado bibliográfico, los recursos empleados para la integración explícita de la información y el tratamiento de los referentes buscaron determinar los rasgos inherentes al uso de fuentes ajenas en la escritura académica, por parte de estudiantes recién ingresados a la educación superior. Estos resultados manifiestan vacilaciones en el uso de las formas, que implicarían un nivel básico en el ingreso a la cultura escrita académica. No obstante, es complejo determinar alguna comparación implicada con los rasgos detectados en las comunidades discursivas, puesto que, tal como se señaló anteriormente, se desconocen los rasgos constitutivos de la transmisión del conocimiento ajeno de las culturas disciplinarias en español.

4. El análisis sobre la presencia de plagio académico según la definición de Pecorari (2008) (60,5% en LLyLH y 88,5% en PCI) resultó fundamental para determinar que las características lingüísticas que lo definen deben variar, de acuerdo con el nivel universitario en el que se encuentre el estudiante. En este sentido, no se debe olvidar que cualquier tarea de escritura y, por tanto, cualquier caso identificable de plagio, ocurre en un contexto educativo que debe ser considerado como un factor al momento de analizar sus causas y posibles sanciones. Consideramos que en los primeros niveles de enseñanza deben proveerse, gradualmente, los recursos formales que permitan diferenciar la información propia de la que no es. Resulta fundamental fomentar la conciencia relativa al reconocimiento de la información ajena. Para ello sería útil la promoción del uso de citas integrales, en la medida en que en ellas la atribución viene determinada con anterioridad. A pesar de lo anterior, si bien serviría para comenzar con la conciencia sobre la información ajena, podría ocasionar dificultades, en el caso en el que en las disciplinas, la norma implícita defina otro tipo de integración de la información como el modelo mayoritario a seguir. Por otro lado, se requieren criterios claros, identificables y cuantificables que determinen el plagio académico, independientemente de la intención de engañar de los estudiantes (cf. Larkham y Manns, 2002). De esta manera, postulamos que los rasgos más claros son los planteados por Pecorari (2008), los que, con todo, deben ser matizados según la cantidad de tiempo que el estudiante lleva en la universidad. Además, los docentes

deberían guiar lo que se entiende por *conocimiento común* en cada disciplina, con el propósito de ayudar el ingreso a las comunidades, facilitando los procesos de atribución de la información.

5. Los datos indicaron que se presentó una correlación entre la representación de la comprensión de múltiples fuentes, derivada con LSA, y la presencia de plagio académico: los plagios se ubican, principalmente, en los grupos de documentos que presentan modelos situacionales menos adecuados. Asimismo, existe una correlación entre la aproximación a los modelos situacionales derivados de LSA y la cantidad de citas utilizadas: mientras más adecuado es el modelo situacional, mayor cantidad de citas se encuentran. Dichos resultados constituyen un aporte a la caracterización sobre la relación entre comprensión de lectura y habilidades de escritura, en la medida en que, en efecto, escribir siempre es escribir sobre algo, y por lo general es sobre algo que se ha leído (Carlino, 2005), por lo que la escritura académica evidenciaría una relación directa y necesaria con la comprensión lectora. Las deficiencias escriturales no se identificarían tan solo por las dificultades formales que presentarían los textos, sino que, además, implicaría la construcción de un modelo situacional poco adecuado, derivado de la comprensión de textos. Dado esto, se reafirma la necesidad de que en la alfabetización académica la expresión escrita no se centre, tan solo, en las características textuales y formales que forman a un texto cohesivo, coherente y adecuado a la situación comunicativa, sino que debe contemplar un proceso reflexivo de comprensión de textos que servirán de *input* a su desarrollo posterior, considerando que la *función* que tiene la escritura para los estudiantes es, por lo general, dar cuenta del aprendizaje de los contenidos de alguna asignatura, a través de géneros discursivos diferentes a los que se emplean, efectivamente, en sus disciplinas.

6. Para finalizar, recalcamos lo exploratorio de este trabajo y las limitaciones impuestas por el método de recolección de información y las características de las asignaturas desde donde se obtuvieron los datos. En este sentido, creemos que esta investigación debería ampliarse y considerar lo que sucede en niveles universitarios superiores, con tareas de escritura diferentes y con fuentes no precisadas, incentivando el uso de recursos

provenientes de a *Web*, lo que permitiría dar luces sobre el tipo de documentos revisados por los estudiantes, en la era de la sociedad de la información.

## NOTAS

<sup>1</sup> Esta acepción se sigue manteniendo en el español de América, según asegura el DRAE

<sup>2</sup> El criterio para categorizar como *prestigiosas* fue el Ranking Arwu (2010) sobre la calidad de la educación superior en el mundo.

<sup>3</sup> Se consideran, exclusivamente, los reglamentos generales de pre-grado, sin hacer mención a los reglamentos internos que pueda tener cada facultad o departamento.

<sup>4</sup> Beer (2000) señala que los sistemas dinámicos son objetos matemáticos que, carentes de ambigüedades, describen cómo evoluciona el estado de algún sistema a través del tiempo.

<sup>5</sup> De *Comprehension*, W. Kintsch, 1998, p. 294.

<sup>6</sup> Definidas como el conocimiento procedimental que tienen las personas acerca de la comprensión de discursos.

<sup>7</sup> En psicología cognitiva, los esquemas se refieren a las “estructuras de carácter jerárquico mediante las cuales el conocimiento se representa y se organiza en la memoria” (García, 2006, p. 283).

<sup>8</sup> Entendido como la parte del contexto sobre el que se interpreta un discurso (Kintsch, 1996, p. 71).

<sup>9</sup> A partir de esta perspectiva, se igualan las nociones de *conceptos* y *proposiciones*, y se pueden tratar indistintamente (Kintsch, 1988)

<sup>10</sup> Adaptado de El rol del conocimiento en la comprensión del discurso: Un modelo de construcción-integración. W. Kintsch, 1996, p. 87.

<sup>11</sup> En cuanto a su denominación, la literatura indica que no existe consenso, mientras unos lo denominan modelos situacionales otros prefieren el término modelos mentales para referirse al mismo tipo de representación mental (cf. Vega, Díaz y León, 1999; Zwaan y Radvansky, 1998).

<sup>12</sup> Definida por la autora como aquellas tareas que tienen como objetivo que la representación elaborada durante el procesamiento presente una estabilidad alta en la memoria.

<sup>13</sup> Existen diferentes medidas que permiten establecer la similitud de significados. En primer lugar, el largo del vector; luego, el *coseno* aplicado al ángulo entre los vectores; y,



por último, la distancia entre vectores. Dentro de éstas, la medida más empleadas es la distancia *coseno*, puesto que es más adecuada cuando se tiene alguna idea previa sobre la similitud semántica entre los documentos que se compararán (Gutiérrez, 2005).

<sup>14</sup> Para Peirce, los *tipos* (types) se relacionan, generalmente, con una entidad abstracta, en cambio, las ocurrencias (*tokens*) son concretas y particulares (Wetzel, 2011).

<sup>15</sup> De “Mathematical Foundations Behind Latent Semantic Analysis”, por D. Martin y M. Berry, 2007, p. 40.

<sup>16</sup> Basado en “Análisis Semántico Latente: ¿Una teoría psicológica del significado?”, R.M. Gutiérrez, 2005.

<sup>17</sup> Basado en “La situación retórica: Su importancia en el aprendizaje y en la enseñanza de la producción escrita”, por R. Benítez, 2000.

<sup>18</sup> Los ejemplos precedidos de asterisco indican realizaciones incorrectas.

<sup>19</sup> La elección de estas regiones responde a que juntas concentran el 45% de la población nacional (Molina et al., 2010).

<sup>20</sup> No deben confundirse los términos *Web* e *Internet*. Mientras el último se relaciona con la red física que conecta diferentes computadores, la *Web* se identifica con la estructura lógica de la información que se aloja en la red física (Gutiérrez, 2008, p. 10).

<sup>21</sup> De “Minimizing plagiarism by redesigning the learning environment and assessment”, por Bhattacharya, M. y Jorgensen, L. 2008, p. 204.

<sup>22</sup> Adaptado de In Other (People’s) Words: Plagiarism by university students - literature and lessons, por Park, C., 2003, p. 475.

<sup>23</sup> Agradecemos a los profesores Cristián Lagos, del Departamento de Lingüística de la Universidad de Chile, y Héctor Augusto, de la Facultad de Ciencias Físicas y Matemáticas de la misma casa de estudios, por su desinteresado apoyo y cooperación en la recolección de los documentos.

<sup>24</sup> Tanto la extracción de *stopwords* como el *stemming* son operaciones necesarias para cualquier análisis de lenguaje natural y no son exclusivos de LSA (cf. Bécue-Bertaut, 2010).

<sup>25</sup> Cabe recordar que la construcción de las matrices puede realizarse por medio de cualquier unidad de análisis: palabra, grupos de palabras, proposiciones, etc.

<sup>26</sup> Realizado gracias a la herramienta ®*Docode Lite*, elaborada con el objetivo de determinar el plagio académico proveniente de la *Web*.

<sup>27</sup> La autora considera a la variante canónica de discurso directo como la presencia de *verbum dicendi*, comillas y dos puntos (Brunetti, 2009, p. 29).

## 6. REFERENCIAS BIBLIOGRÁFICAS

- American Psychological Association. (2001). *Manual de estilo de publicaciones de la American Psychological Association*. México D.F.: Manual Moderno.
- Antoniou, G. y Van Harmeleen, F. (2002). *A Semantic Web Primer*. Massachusetts: MIT Press.
- Bazerman, Ch. y Paradis, J. (2004[1991]). *Textual Dynamics of the Professions Historical and Contemporary Studies of Writing in Professional Communities*. Madison: University of Wisconsin Press.
- Bazerman, Ch. (2004). "Intertextuality: How Texts Rely on Other Texts". En Bazerman, Ch. y Prior, P. (Eds.), *What writing does and how it does it: an introduction to analyzing texts and Textual Practices*. New Jersey: Lawrence Erlbaum Associates.
- Bécue-Bertaut, M. (2010). *Minería de textos. Aplicación a preguntas abiertas en encuestas*. Madrid: La Muralla.
- Beer, R. (2000). Dynamical approaches to cognitive science. *Trends in Cognitive Science* 4(3), 91-99.
- Belinchón, M., Igoa, J., Rivière, A. (2000). *Psicología del lenguaje. Investigación y teoría*. Madrid: Trotta.
- Benítez, R. (2000). La situación retórica: Su importancia en el aprendizaje y en la enseñanza de la producción escrita. *Signos* 33(48), 49-67.
- Berkenkotter, C., Huckin, T. y Ackerman, J. (2004[1991]). "Social Context and Socially Constructed Texts: The Initiation of a Graduate Student into a Writing Research Community". En Bazerman, Ch. y Paradis, J. *Textual Dynamics of the Professions Historical and Contemporary Studies of Writing in Professional Communities* (191-215). Madison: University of Wisconsin Press.

- Bhattacharya, M. y Jorgensen, L. (2008). "Minimizing plagiarism by redesigning the learning environment and assessment". En Roberts, T. (Ed.), *Student Plagiarism in an Online World: Problems and Solutions* (pp. 194-207). USA: Information Science Reference.
- Bravo-Márquez, F., L'Huillier, G., Moya, P., Ríos, S., Velásquez, J. (2011). *An Automatic Text Comprehension Classifier Based on Mental Models and Latent Semantic Features*. Manuscrito presentado para su publicación.
- Brunetti, P. (2009). *El discurso referido: Formas canónicas y no canónicas de citación en la prensa diaria. Aspectos teóricos y didácticos*. Córdoba, Argentina: Comunicarte.
- Carlino, P. (2003). Alfabetización académica: un cambio necesario, algunas alternativas posibles. *Educere* 20, 409-420.
- Carlino, P. (2005). *Escribir, leer y aprender en la universidad: una introducción a la alfabetización académica*. México: Fondo de cultura económica.
- Carreiras, M. (1997). *Descubriendo y procesando el lenguaje*. Madrid: Trotta.
- Cerdán, R., (2006). *Integration Information Processes from Multiple Documents*. Tesis para optar al Grado de Doctor en Psicología, Universidad de Valencia, Valencia.
- Češka, Z. (2009). *Automatic Plagiarism Detection Based on Latent Semantic Analysis*. Tesis para optar al grado de Doctor en Ingeniería y Ciencias de la Computación, Universidad de West Bohemia, Pilsen, República Checa.
- Colón, G. (1992). Reflejos cultos de plagium y plagiarius en algunas lenguas europeas. *Voces* III, 9-19.
- De Beaugrande, R. y Dressler, W. (1981). *Introduction to text linguistics*. Londres: Longman.

- Foltz, P., Britt, M.A., Perfetti, Ch. (1996). Reasoning from multiple texts: An automatic analysis of readers' situation models. En Cottrell, G. (Ed.), *Proceedings of the 18th Annual Cognitive Science Conference* (pp. 110-115). New Jersey: Lawrence Erlbaum.
- García, J. (2006). *Lectura y conocimiento*. Barcelona: Paidós.
- Graesser, A., Gernsbacher, M. y Goldman, S. (2000). "Cognición". En Van Dijk, T. (Comp.), *El discurso como estructura y proceso* (pp. 417-451). Barcelona: Gedisa.
- Green, S. (2002). Plagiarism, Norms, and the Limits of Theft Law: Some observations on the Use of Criminal Sanctions in Enforcing Intellectual Property Rights. *Hastings Law Journal* 54(1), 167-242.
- Guerrero, G. (2010). *Modelo de fusión de datos por score utilizando teoría de la información para integración de sistemas de similitud de documentos*. Tesis para optar al grado de Magíster en Gestión de Operaciones, Universidad de Chile, Chile.
- Gutiérrez, R.M. (2005). Análisis Semántico Latente: ¿Teoría psicológica del significado? *Signos* 38(59), 303-323.
- Gutiérrez, C. (2008). "La Web como espacio de información universal". En Centro de Investigación de la Web, *Cómo funciona la Web* (pp.9-21). Recuperado el 10 de julio de 2010 de [www.ciw.cl](http://www.ciw.cl).
- Haviland, C. y Mullin, J. (2009). *Who Owns this Text? Plagiarism, Authorship and Disciplinary Cultures*. Logan, Utah: Utah State University Press.
- Hyland, K. (1999). Academic attribution: Citation and the construction of disciplinary knowledge. *Applied Linguistics*, 20(3), 341-367.
- Hyland, K. (2000). *Disciplinary Discourses: Social interactions in academic Writing*. Harlow, UK: Longman.
- Hyland, K. (2002). *Teaching and Researching Writing*. Harlow, UK: Longman.

- Hyland, K. (2005). Stance and engagement: a model of interaction in academic discourse. *Discourse Studies* 7(2), 173-192.
- Kintsch, W. (1988). The role of knowledge in discourse comprehension: a construction-integration model. *Psychological Review* 95 (2), pp. 162-182.
- Kintsch, W. (1994). Text Comprehension, Memory, and Learning. *American Psychologist*, 49(4), 294-303.
- Kintsch, W. (1996). "El rol del conocimiento en la comprensión del discurso: Un modelo de construcción-integración". En *Textos en contexto 2: Los procesos de lectura y escritura* (pp. 78-138). Buenos Aires: Asociación Internacional de Lectura.
- Kintsch, W. (1998). *Comprehension: a paradigm for cognition*. Cambridge: Cambridge University Press.
- Kintsch, W. (2002). "On the notions of theme and topic in psychological process models of text comprehension". En Louwerse, M. y Van Peer, W. (Eds.), *Thematics: Interdisciplinary Studies* (pp. 157-160). Amsterdam: Benjamins.
- Kintsch, W. (2007). "Meaning in Context". En Landauer, T., McNamara, D., Dennis, S. y Kintsch, W. (Eds.), *Handbook of Latent Semantic Analysis* (pp. 89-106). New Jersey: Lawrence Erlbaum Associates.
- Kintsch, W. y Kintsch, E. (2005). "Comprehension". En Paris, S. y Stahl, S. (Eds.), *Children's Reading Comprehension and Assessment* (pp. 71-92). New Jersey: Lawrence Erlbaum Associates.
- Kintsch, W. y Van Dijk, T. (1978). Toward a model of text comprehension. *Psychological Review* 85 (5), pp. 363-394.
- Kintsch, W. y Van Dijk, T. (1983). *Strategies of Discourse Comprehension*. New York: Academic Press.

- Lancaster, T. y Culwin, F. (2004). A Comparison of Source Code Plagiarism Detection Engines. *Computer Science Education* 14(2), 101-117.
- Landauer, T., Foltz P. y Laham, D. (1998). Introduction to Latent Semantic Analysis. *Discourse Processes* 25, 259-284.
- Larkham, P. y Manns, S. (2002). Plagiarism and its Treatment in Higher Education. *Journal of Further and Higher Education* 26(4), 339-349.
- Lemaire, B., Denhière, G., Bellissens, C., Jhean-Larose, S. (2006). A computational model for simulating text comprehension. *Behaviour Research Methods* 38(4), 628-637.
- Lemaire, B., Denhière, G., Bellissens, C., Jhean-Larose, S. (2007). "A semantic space for modeling children's semantic memory". En Landauer, T., McNamara, D., Dennis, S. y Kintsch, W. (Eds.), *Handbook of Latent Semantic Analysis* (pp. 143-165). New Jersey: Lawrence Erlbaum Associates.
- Ley N°17.336 sobre Propiedad intelectual. (2004). Recuperada el 10 de noviembre de 2010 desde [www.bcn.cl](http://www.bcn.cl).
- Martin, D. y Berry, M. (2007). "Mathematical Foundations Behind Latent Semantic Analysis". En Landauer, T., McNamara, D., Dennis, S. y Kintsch, W. (Eds.), *Handbook of Latent Semantic Analysis* (pp. 35-56). New Jersey: Lawrence Erlbaum Associates.
- Maldonado, C. (1999). "Discurso directo y discurso indirecto". En Bosque, I. y Demonte, V. (Dirs.), *Gramática descriptiva de la lengua española* (pp. 3550-3595). Madrid: Espasa.
- Merriam-Webster Incorporated. (2010). *Merriam Webster Online Dictionary*. Recuperado el 20 de octubre de 2010 de <http://www.merriam-webster.com>.
- Molina, F., Velásquez, J.D., Ríos, S., Calfucoy, P., y Cociña, M. (2011). El fenómeno del Plagio en Documentos Digitales: Un Análisis de la Situación Actual en el Sistema Educativo. *Revista de Ingeniería de Sistemas*, XXV (en prensa).

- Moore Howard, R. (2007). Understanding 'Internet Plagiarism'. *Computers and Composition* 24, 3-15.
- Moya, P., L'Huillier, G., Velásquez, J. y Ríos, S. (2010). *Análisis automatizado de la similitud entre documentos universitarios: una aproximación a la detección del plagio académico y al uso de fuentes de información*. Manuscrito presentado para su publicación.
- Moyano, E. (2010). Escritura académica a lo largo de la carrera. Un programa institucional. *Signos* 43(74), 465-488.
- Oxford University Press. (2010). *Oxford Dictionaries Online*. Recuperado el 20 de octubre de 2010 de <http://oxforddictionaries.com>.
- Park, C. (2003). In Other (People's) Words: plagiarism by university students—literature and lessons. *Assessment & Evaluation in Higher Education* 28 (5), 471-488.
- Pecorari, D. (2003). Good and original: Plagiarism and patchwriting in academic second-language writing. *Journal of Second Language Writing*, 12, 317-345.
- Pecorari, D. (2008). *Academic Writing and Plagiarism: A Linguistic Analysis*. Londres: Continuum.
- Perfetti, Ch., Rouet, J. y Britt, M.A. (1999). Toward a Theory of Documents Representation. En Van Oostendorp, H. y S. Goldman (Eds.), *The Construction of Mental Representations During Reading* (pp. 99-122). London: LEA.
- Pontificia Universidad Católica de Chile. (2008). *Reglamento del alumno de pregrado*. Recuperado el 20 de diciembre de 2010 desde [www.puc.cl](http://www.puc.cl).
- Pupovac, V., Bilič-Zulle, L. y Petrovečki, M. (2008). On academic plagiarism in Europe. An analytical approach based on four studies. En Comas, R., y Sureda, J. (Coords.), *Academic Cyberplagiarism* [dossier online], *Digithum* 10. Recuperado el 12 de junio de 2010.



- Real Academia de la Lengua Española. (1973). *Esbozo de una nueva gramática de la lengua española*. Madrid: Espasa.
- Real Academia de la Lengua Española. (2001). *Diccionario de la lengua española*. Madrid: Espasa.
- Real Academia de la Lengua Española. (2010). *Nueva gramática de la lengua española (Manual)*. Madrid: Espasa.
- Reyes, G. (1995). *Los procedimientos de cita: el estilo directo y estilo indirecto*. Madrid: Arco Libros.
- Rouet, J.-F. (2006). *The skills of document use: From text comprehension to Webbased learning*. Mahwah, NJ: Erlbaum.
- Rouet, J.-F., Favart, M., Britt, M. A., y Perfetti, C. A. (1997). Studying and using multiple documents in history: Effects of discipline expertise. *Cognition and Instruction*, 15, 85-106.
- Rus, V., Lintean, M. y Azevedo, R. (2009). Automatic Detection of Student Mental Models During Prior Knowledge Activation in MetaTutor. En Barnes, T., Desmarais, M., Romero, C., y Ventura, S. (Eds.). *Educational Data Mining 2009: 2nd International Conference on Educational Data Mining, Proceedings*.
- ShanghaiRanking Consultancy. (2010). *Academic Ranking of World Universities (ARWU)*. Recuperado el 15 de enero de 2011 desde [www.arwu.org](http://www.arwu.org).
- Silvestri, A. (1999). Dificultades iniciales de escolares adolescentes en el procesamiento de fuentes múltiples expositivo-narrativas. *Lenguas Modernas*, 26-27, 53-64.
- Solé, I. (s.f.). *La lectura como transformación del pensamiento*. Recuperado el 20 de diciembre de 2009 desde <http://sol-e.com/plec/documentos>.

- Soto, G. (2002). La argumentación en artículos científicos escritos en español. En *Actas del Congreso Internacional La Argumentación: Lingüística, Retórica, Lógica, Pedagogía*, Universidad de Buenos Aires, Buenos Aires.
- Soto, G. (2009). Intertextualidad explícita en textos académicos de estudiantes universitarios. Un estudio exploratorio. *Literatura y Lingüística* 20, 141-157.
- Sutherland-Smith, W. (2008). *Plagiarism, the Internet, and student learning: improving academic integrity*. New York: Routledge.
- Swales, J. (1990). *Genre Analysis: English in academic and research settings*. Cambridge: Cambridge University Press.
- TEI Consortium. (2010). *Guidelines for Electronic Text Encoding and Interchange*. Recuperado el 30 de agosto de 2010 desde <http://www.tei-c.org/P5/>.
- Terreros y Pandó, E. (1987[1786]). *Diccionario castellano con las voces de ciencias y artes*. Madrid: Arco Libros.
- Turell, M.T. (2005). "El plagio en la traducción literaria". En Turell, M.T. (ed.), *Lingüística forense, lengua y derecho. Conceptos, métodos y aplicaciones* (pp. 275-288). Barcelona: Instituto Universitario de Lingüística Aplicada.
- Universidad de Chile. (1993). *Reglamento general de estudiantes universitarios de pregrado*. Recuperado el 20 de diciembre de 2010 desde [www.uchile.cl](http://www.uchile.cl).
- Van Dijk, T. (1992). *La ciencia del texto*. Barcelona: Paidós.
- Vega, A. (2011). *Comprensión de múltiples textos expositivos: relaciones entre conocimiento previo y autorregulación*. Tesis para optar al grado de Doctora en Psicología de la Educación. Universidad Ramon Llull, Barcelona.
- Vega, M. de, Díaz, J. M., y León, I. (1999). "Procesamiento del discurso". En Vega, M. de Cuetos, F. (Eds.), *Psicolingüística del español* (pp. 271-289). Madrid: Trotta.

- Venegas, R. (2005). *Las relaciones léxico-semánticas en artículos de investigación científica: una aproximación desde el análisis semántico latente*. Tesis para optar al grado de Doctor en Lingüística, Pontificia Universidad Católica de Valparaíso, Valparaíso.
- Wetzel, L. (2011). "Types and Tokens". En Zalta, E. (ed.), *The Stanford Encyclopedia of Philosophy*. Recuperado el 12 de marzo de 2011 desde <http://plato.stanford.edu/archives/spr2011/entries/types-tokens/>.
- Wiley, J. y Voss, J. F. (1999). Constructing arguments from multiple sources: Tasks that promote understanding and not just memory for text. *Journal of Educational Psychology*, 91, 301-311.
- Zwaan, R. y Radvansky, G. (1998). Situation Models in Language Comprehension and Memory. *Psychological Bulletin*, 123 (2), 162-185.

## 7. ANEXOS

### Anexo 1: Cuadro de detalle de justificación para el uso del plagio

Justificación	Frecuencia de copia	
	Una vez' o 'nunca'	Más de una vez ('más de una' o 'Muchas' veces)
Permite ahorrar tiempo	31,4%	55,3%
El trabajo queda mejor que si lo hago por mi mismo	32,3%	41,8%
Me es difícil escribir o sintetizar información	10,2%	22,2%
No sé hacerlo de otra manera	4,2%	6,2%
Es lo que me piden	10,4%	15,2%
A la gente que copia le va mejor	6,4%	11,4%
Lo hago porque todos lo hacen	3,6%	8,1%
Nunca creí que era un problema	7,4%	15,5%
Creo que al profesor no le importa	6,9%	18,7%
Creo que el profesor no se da cuenta	7,3%	17,5%
No me queda tiempo para hacerlo de otra manera	15,9%	29,9%

*Nota 1.* Molina et al. (2011).

Anexo 2: Detalle bibliográfico de los documentos de base

Carrera	Documento
PCI	Bilbao, G., Fuertes, J. y Guibert, J.M. (2006). “Presentación” e “Introducción”. En <i>Ética para ingenieros</i> (9-19). Bilbao: Desclée de Bouwer.
	Gallegos, H. (2004). <i>La ética en la ingeniería</i> . Discurso presentado en Lima sin más referencias.
LLyLH	Bunak, V. (1972). Del grito a la palabra. <i>El correo de la UNESCO</i> , agosto-septiembre, 57-58, 69-70.
	Bouknak, V. V. (1969). “El origen del lenguaje”. En <i>Los procesos de hominización</i> , Vallois, H., et al. (93-110). México: Grijalbo.

Anexo 3: Estímulos a responder por los estudiantes

Carrera	Encabezado	Estímulo
PCI	Basándose en los textos entregados, responda la siguiente pregunta. La extensión mínima es de una (1) plana y la máxima de dos (2), usando la fuente Times New Roman 12. Utilice en su respuesta algún sistema de referencias bibliográficas.	Explique la naturaleza ética de la ingeniería. Refiérase en su respuesta a la Ecoingeniería y a las consecuencias del atentado del 11 de marzo en Madrid.
LLyLH		Explique la relación que se puede establecer entre el origen del lenguaje y las ideas generales, conceptos, uso de herramientas y forma de la cabeza

#### Anexo 4: Ejemplo de documento XML

```
<?xml version="1.0" encoding="UTF-8"?>
<document>
<tipo>prueba</tipo>
<autor>
<nombre></nombre>
<carrera>ingenieria</carrera>
</autor>
<asignatura>Introducción a la Ingeniería</asignatura>
<body>
<block>
<pregunta>Explique la naturaleza ética de la ingeniería. Refiérase en su respuesta a las
consecuencias del atentado del 11 de marzo en Madrid</pregunta>
<respuesta>La ética en la Ingeniería es un término un poco subjetivo, que depende del criterio de
cada persona, pero que, sin embargo, es un tema de gran importancia para la sociedad, donde su
presencia en los proyectos, afecta directamente a la sociedad y a la calidad de vida de muchas
personas. Como un ejemplo de esto último, señalamos el atentado del 11 de marzo del 2004 en
Madrid, donde un mal uso de la técnica del ingeniero, provoca un daño hacia una gran cantidad de
personas, demostrando lo que el ser humano es capaz de hacer cuando existe el poder necesario
para realizar y no se tienen límites. Por estos hechos es que se descubre que se necesita llegar a
un acuerdo en cuanto a las responsabilidades y valores que se ejercen sobre la sociedad. En este
contexto, es donde nace la ética, en donde es necesario pensar en las consecuencias de las
creaciones humanas con respecto a la naturaleza y a las repercusiones sociales que tendrán éstas
mismas. Este punto, la ética es la parte fundamental para cuidarnos a nosotros mismos de lo que
somos capaces de hacer, asumiendo la responsabilidad ante el poder que se tiene de seguir
creando. Actualmente, se podría decir que se generan muchos problemas éticos, ya que
últimamente es un tema con mucha relevancia dentro de la sociedad; por lo que también se crean
medidas de formación ética y sanción. Como el mundo en el que vivimos requiere de la
explotación de muchos recursos, aparece lo que se denomina la Eco-ingeniería, que es un modo
más ético de explotación de los recursos naturales que tenemos para el uso de la sociedad. Esto
mismo nace cuando la sociedad toma conciencia de que tiene que cuidar el mundo en el que se
vive para poder seguir viviendo en él y aprovechando sus recursos, y esto mismo es una aplicación
de la ética en la vida cotidiana. Tal como señala Bilbao <cita>Otro tema importante es la
responsabilidad de las personas a cargo de un proyecto de ingeniería en la toma de decisiones, ya
que muchas veces están en juego la vida de los trabajadores y personas cercanas a la fuente de
trabajo.<tipo>directa</tipo></cita> Como conclusión, la ética en la ingeniería es un tema
demasiado importante, como para pasarlo por alto, ya que si bien la ingeniería aporta soluciones a
los problemas cotidianos de las personas, si se ejecuta erróneamente puede ocasionar muchos
problemas a la naturaleza, a la sociedad, al país, o al lugar donde se esté trabajando.</respuesta>
</block>
</body>
</document>
```

Anexo 5: Distancia LSA n-grama 1 y 2, en LLyLH

LLyLH	N-grama 1			N-grama 2		
	Texto 1	Texto 2	Promedio	Est.1	Est. 2	Promedio
<b>Alumno 1</b>	0,48	0,52	0,50	0,25	0,16	0,20
<b>Alumno 2</b>	0,91	0,66	0,79	0,83	0,18	0,51
<b>Alumno 3</b>	0,52	0,60	0,56	0,21	0,14	0,18
<b>Alumno 4</b>	0,56	0,50	0,53	0,32	0,13	0,22
<b>Alumno 5</b>	0,74	0,64	0,69	0,31	0,21	0,26
<b>Alumno 6</b>	0,61	0,60	0,60	0,20	0,26	0,23
<b>Alumno 7</b>	0,62	0,56	0,59	0,45	0,15	0,30
<b>Alumno 8</b>	0,58	0,56	0,57	0,25	0,20	0,23
<b>Alumno 9</b>	0,48	0,51	0,49	0,13	0,14	0,14
<b>Alumno 10</b>	0,44	0,49	0,46	0,23	0,21	0,22
<b>Alumno 11</b>	0,60	0,59	0,60	0,48	0,28	0,38
<b>Alumno 12</b>	0,51	0,65	0,58	0,15	0,29	0,22
<b>Alumno 13</b>	0,54	0,53	0,54	0,24	0,12	0,18
<b>Alumno 14</b>	0,67	0,52	0,60	0,42	0,16	0,29
<b>Alumno 15</b>	0,47	0,44	0,45	0,14	0,09	0,12
<b>Alumno 16</b>	0,59	0,60	0,59	0,26	0,15	0,20
<b>Alumno 17</b>	0,60	0,73	0,66	0,28	0,29	0,28
<b>Alumno 18</b>	0,75	0,64	0,69	0,76	0,23	0,50
<b>Alumno 19</b>	0,71	0,68	0,69	0,54	0,23	0,38
<b>Alumno 20</b>	0,48	0,46	0,47	0,09	0,11	0,10
<b>Alumno 21</b>	0,58	0,62	0,60	0,21	0,25	0,23
<b>Alumno 22</b>	0,59	0,63	0,61	0,24	0,27	0,25
<b>Alumno 23</b>	0,61	0,56	0,59	0,17	0,21	0,19



<b>Alumno 24</b>	0,58	0,55	0,56	0,22	0,43	0,32
<b>Alumno 25</b>	0,56	0,58	0,57	0,34	0,28	0,31
<b>Alumno 26</b>	0,53	0,55	0,54	0,17	0,20	0,19
<b>Alumno 27</b>	0,44	0,49	0,46	0,06	0,12	0,09
<b>Alumno 28</b>	0,47	0,59	0,53	0,08	0,09	0,09
<b>Alumno 29</b>	0,44	0,60	0,52	0,07	0,13	0,10
<b>Alumno 30</b>	0,60	0,56	0,58	0,28	0,21	0,24
<b>Alumno 31</b>	0,64	0,63	0,63	0,21	0,14	0,17
<b>Alumno 32</b>	0,65	0,58	0,61	0,30	0,17	0,23
<b>Alumno 33</b>	0,57	0,67	0,62	0,17	0,18	0,18
<b>Alumno 34</b>	0,71	0,75	0,73	0,41	0,38	0,39
<b>Alumno 35</b>	0,59	0,58	0,58	0,24	0,22	0,23
<b>Alumno 36</b>	0,59	0,62	0,61	0,41	0,40	0,40
<b>Alumno 37</b>	0,63	0,73	0,68	0,30	0,27	0,28
<b>Alumno 38</b>	0,47	0,59	0,53	0,13	0,26	0,20
<b>Alumno 39</b>	0,67	0,61	0,64	0,48	0,26	0,37
<b>Alumno 40</b>	0,51	0,47	0,49	0,14	0,05	0,10
<b>Alumno 41</b>	0,63	0,56	0,59	0,44	0,23	0,33
<b>Alumno 42</b>	0,40	0,43	0,42	0,05	0,11	0,08
<b>Alumno 43</b>	0,64	0,61	0,62	0,36	0,18	0,27

Anexo 6: Distancia LSA n-grama 1 y 2, en PCI

PCI	N-grama 1			N-grama 2		
	Texto 1	Texto 2	Promedio	Texto 1	Texto 2	Promedio
<b>Alumno 1</b>	0,36	0,41	0,39	0,24	0,08	0,16
<b>Alumno 2</b>	0,60	0,49	0,55	0,15	0,20	0,18
<b>Alumno 3</b>	0,60	0,58	0,59	0,18	0,10	0,14
<b>Alumno 4</b>	0,63	0,50	0,57	0,10	0,12	0,11
<b>Alumno 5</b>	0,46	0,47	0,46	0,11	0,16	0,14
<b>Alumno 6</b>	0,53	0,32	0,43	0,03	0,01	0,02
<b>Alumno 7</b>	0,55	0,58	0,56	0,12	0,21	0,17
<b>Alumno 8</b>	0,60	0,50	0,55	0,17	0,13	0,15
<b>Alumno 9</b>	0,71	0,62	0,66	0,13	0,25	0,19
<b>Alumno 10</b>	0,44	0,40	0,42	0,06	0,01	0,04
<b>Alumno 11</b>	0,47	0,36	0,42	0,07	0,13	0,10
<b>Alumno 12</b>	0,35	0,26	0,30	0,01	0,03	0,02
<b>Alumno 13</b>	0,60	0,46	0,53	0,06	0,04	0,05
<b>Alumno 14</b>	0,57	0,50	0,54	0,12	0,12	0,12
<b>Alumno 15</b>	0,47	0,36	0,42	0,09	0,01	0,05
<b>Alumno 16</b>	0,62	0,58	0,60	0,16	0,16	0,16
<b>Alumno 17</b>	0,69	0,49	0,59	0,29	0,11	0,20
<b>Alumno 18</b>	0,51	0,42	0,46	0,15	0,05	0,10
<b>Alumno 19</b>	0,46	0,49	0,47	0,09	0,17	0,13
<b>Alumno 20</b>	0,63	0,50	0,56	0,21	0,09	0,15
<b>Alumno 21</b>	0,49	0,49	0,49	0,10	0,10	0,10
<b>Alumno 22</b>	0,57	0,54	0,56	0,10	0,10	0,10
<b>Alumno 23</b>	0,40	0,45	0,43	0,05	0,17	0,11
<b>Alumno 24</b>	0,35	0,44	0,39	0,08	0,07	0,08
<b>Alumno 25</b>	0,64	0,51	0,57	0,11	0,04	0,07

<b>Alumno 26</b>	0,67	0,55	0,61	0,14	0,19	0,17
<b>Alumno 27</b>	0,48	0,45	0,47	0,09	0,01	0,05
<b>Alumno 28</b>	0,56	0,46	0,51	0,18	0,07	0,13
<b>Alumno 29</b>	0,73	0,63	0,68	0,19	0,11	0,15
<b>Alumno 30</b>	0,59	0,51	0,55	0,17	0,14	0,16
<b>Alumno 31</b>	0,55	0,45	0,50	0,03	0,01	0,02
<b>Alumno 32</b>	0,61	0,58	0,59	0,20	0,19	0,19
<b>Alumno 33</b>	0,71	0,53	0,62	0,17	0,11	0,14
<b>Alumno 34</b>	0,42	0,51	0,47	0,11	0,11	0,11
<b>Alumno 35</b>	0,62	0,53	0,58	0,10	0,28	0,19
<b>Alumno 36</b>	0,69	0,51	0,60	0,18	0,05	0,11
<b>Alumno 37</b>	0,48	0,63	0,56	0,02	0,31	0,16
<b>Alumno 38</b>	0,59	0,50	0,55	0,09	0,28	0,18
<b>Alumno 39</b>	0,52	0,40	0,46	0,07	0,13	0,10
<b>Alumno 40</b>	0,54	0,46	0,50	0,02	0,12	0,07
<b>Alumno 41</b>	0,48	0,67	0,58	0,08	0,22	0,15
<b>Alumno 42</b>	0,46	0,27	0,37	0,18	0,01	0,10
<b>Alumno 43</b>	0,60	0,54	0,57	0,14	0,03	0,08
<b>Alumno 44</b>	0,61	0,31	0,46	0,21	0,01	0,11
<b>Alumno 45</b>	0,52	0,53	0,53	0,13	0,09	0,11
<b>Alumno 46</b>	0,59	0,66	0,63	0,18	0,15	0,17
<b>Alumno 47</b>	0,57	0,44	0,50	0,03	0,04	0,04
<b>Alumno 48</b>	0,55	0,48	0,51	0,21	0,08	0,14
<b>Alumno 49</b>	0,65	0,56	0,60	0,03	0,01	0,02
<b>Alumno 50</b>	0,57	0,46	0,52	0,12	0,16	0,14
<b>Alumno 51</b>	0,57	0,50	0,53	0,08	0,08	0,08
<b>Alumno 52</b>	0,52	0,39	0,46	0,10	0,05	0,07
<b>Alumno 53</b>	0,60	0,55	0,57	0,23	0,24	0,23
<b>Alumno 54</b>	0,62	0,57	0,59	0,05	0,01	0,03

<b>Alumno 55</b>	0,62	0,39	0,50	0,25	0,01	0,13
<b>Alumno 56</b>	0,57	0,58	0,57	0,17	0,03	0,10
<b>Alumno 57</b>	0,55	0,39	0,47	0,13	0,01	0,07
<b>Alumno 58</b>	0,72	0,52	0,62	0,13	0,08	0,11
<b>Alumno 59</b>	0,65	0,44	0,54	0,22	0,02	0,12
<b>Alumno 60</b>	0,55	0,55	0,55	0,11	0,06	0,08
<b>Alumno 61</b>	0,55	0,52	0,53	0,05	0,15	0,10

Anexo 7: Detalle de cantidad de citas por sujeto según la taxonomía de Swales (1990) ampliada en LLyLH

	<b>Integrada con referencia</b>	<b>Integrada con referencia incompleta</b>	<b>Integrada sin referencia</b>	<b>No integrada con referencia</b>	<b>No integrada con referencia incompleta</b>	<b>No integrada sin referencia</b>	<b>Total citas</b>
Alumno 1	1	0	0	0	0	0	1
Alumno 2	2	0	0	5	0	0	7
Alumno 3	0	0	0	0	0	1	1
Alumno 4	0	0	0	0	0	1	1
Alumno 5	0	0	0	0	2	0	2
Alumno 6	0	0	9	0	0	0	9
Alumno 7	0	0	0	0	0	0	0
Alumno 8	0	0	0	2	0	0	2
Alumno 9	0	0	0	0	0	0	0
Alumno 10	1	0	0	0	0	0	1
Alumno 11	0	0	0	2	0	0	2
Alumno 12	0	0	0	0	0		0
Alumno 13	0	0	0	1	1	0	2
Alumno 14	0	0	0	1	0	0	1
Alumno 15	0	0	0	0	0	0	0
Alumno 16	0	0	0	3	0	0	3
Alumno 17	2	1	2	0	0	4	9
Alumno 18	0	0	0	0	1	1	2
Alumno 19	0	0	4	0	0	0	4
Alumno 20	0	0	0	1	0	0	1
Alumno 21	0	0	0	1	0	0	1
Alumno 22	6	0	0	0	0	0	6
Alumno 23	0	0	0	3	0	0	3

Alumno 24	0	0	0	2	0	0	2
Alumno 25	0	0	2	0	0	0	2
Alumno 26	2	0	4	0	0	0	6
Alumno 27	0	0	0	0	0	0	0
Alumno 28	0	0	0	0	0	0	0
Alumno 29	0	0	0	0	0	0	0
Alumno 30	0	0	0	4	0	0	4
Alumno 31	0	0	5	0	0	0	5
Alumno 32	0	0	3	0	0	0	3
Alumno 33	0	0	0	0	0	0	0
Alumno 34	0	0	0	0	0	0	0
Alumno 35	0	0	0	0	0	0	0
Alumno 36	0	0	0	2	0	0	2
Alumno 37	0	0	1	4	0	0	5
Alumno 38	1	0	0	3	0	0	4
Alumno 39	0	0	1	0	0	0	1
Alumno 40	0	0	0	0	0	0	0
Alumno 41	0	0	0	2	0	0	2
Alumno 42	0	0	1	0	0	0	1
Alumno 43	3	0	0	0	0	0	3
Total	18	1	32	36	4	7	98

Anexo 8: Detalle de cantidad de citas por sujeto según la taxonomía de Swales (1990) ampliada en LLyLH

	<b>Integrada con referencia</b>	<b>Integrada con referencia incompleta</b>	<b>Integrada sin referencia</b>	<b>No integrada con referencia</b>	<b>No integrada con referencia incompleta</b>	<b>No integrada sin referencia</b>	<b>Total citas</b>
Alumno 1	0	0	0	0	0	0	0
Alumno 2	0	0	2	0	0	0	2
Alumno 3	0	0	0	0	4	0	4
Alumno 4	0	1	1	0	0	0	2
Alumno 5	0	0	0	0	0	0	0
Alumno 6	0	0	0	0	0	0	0
Alumno 7	0	0	2	0	1	0	3
Alumno 8	0	0	0	0	1	0	1
Alumno 9	0	0	1	2	0	0	3
Alumno 10	0	0	0	0	0	0	0
Alumno 11	0	0	0	0	0	0	0
Alumno 12	0	0	0	0	0	0	0
Alumno 13	0	0	0	1	0	0	1
Alumno 14	0	0	3	0	0	0	3
Alumno 15	0	0	2	0	0	0	2
Alumno 16	0	0	4	0	0	0	4
Alumno 17	0	0	4	0	0	0	4
Alumno 18	0	0	4	0	0	1	5
Alumno 19	0	0	0	0	2	0	2
Alumno 20	0	0	3	1	0	0	4
Alumno 21	0	0	1	2	0	0	3
Alumno 22	0	0	0	1	0	0	1
Alumno 23	0	1	3	0	0	0	4

Alumno 24	0	0	0	0	0	0	0
Alumno 25	0	0	1	0	1	0	2
Alumno 26	0	0	0	5	0	0	5
Alumno 27	0	0	0	0	0	0	0
Alumno 28	2	0	0	2	0	0	4
Alumno 29	0	0	1	6	0	0	7
Alumno 30	2	0	0	0	0	0	2
Alumno 31	0	0	1	0	0	0	1
Alumno 32	0	0	0	2	0	0	2
Alumno 33	0	0	1	0	0	0	1
Alumno 34	0	0	5	0	0	0	5
Alumno 35	0	0	7	0		0	7
Alumno 36	0	0	3	0	0	0	3
Alumno 37	0	0	0	0	0	0	0
Alumno 38	0	0	0	0	0	0	0
Alumno 39	0	0	0	0	0	0	0
Alumno 40	0	0	3	0	0	0	3
Alumno 41	0	0	1	0	0	1	2
Alumno 42	0	0	3	0	0	0	3
Alumno 43	0	0	2	0	0	0	2
Alumno 44	0	0	2	0	0	0	2
Alumno 45	0	0	1	0	0	0	1
Alumno 46	0	0	2	0	0	0	2
Alumno 47	0	0	1	0	0	0	1
Alumno 48	0	0	2	0	0	1	3
Alumno 49	0	0	1	0	0	1	2
Alumno 50	0	0	0	0	0	0	0
Alumno 51	0	0	0	0	0	0	0
Alumno 52	0	0	0	0	0	0	0



Alumno 53	2	0	0	0	0	0	2
Alumno 54	0	0	2	0	0	0	2
Alumno 55	0	0	0	0	1	0	1
Alumno 56	0	0	1	0	0	0	1
Alumno 57	0	0	3	0	0	0	3
Alumno 58	0	0	3	0	0	0	3
Alumno 59	0	0	4	0	0	0	4
Alumno 60	0	0	1	0	0	0	1
Alumno 61	0	0	0	0	0	1	1
Total	6	2	81	22	10	5	126

---

