Discurso informal matemático: enfoques, métodos, técnicas y tendencias

Informal mathematical discourse: approaches, methods, techniques and tendencies

Raúl Ernesto Gutiérrez de Piñerez Reyes[1] Ph.D. & Juan Francisco Díaz Frías. Ph.D.
1. Escuela de Ingeniería de Sistemas y Computación, Universidad del Valle, Colombia
raul.gutierrez@correounivalle.edu.co, juanfco.diaz@correounivalle.edu.co

Recibido para revisión 02 de marzo de 2011, aceptado 28 de junio de 2011, versión final 05 de julio de 2011

***Resumen*— El Discurso Informal Matemático (DIM) se caracteriza por la mezcla entre lenguaje natural y las expresiones simbólicas que se presentan en el contexto de los libros de textos, en las publicaciones matemáticas y en las demostraciones matemáticas. En la actualidad el discurso informal matemático está siendo abordado desde tres grandes áreas de trabajo; una primera, la asistencia y chequeo automático de teoremas; una segunda, la automatización de los fenómenos lingüísticos que se presentan en las demostraciones matemáticas; y una tercera, que tiene que ver con la gestión del conocimiento matemático sobre grandes corpus de textos matemáticos. En este artículo se exponen los métodos, técnicas y tendencias sobre el DIM con el fin de mostrar la aplicabilidad de diversas teorías lingüístico-computacionales de manera que sirvan de punto de partida para el desarrollo de proyectos futuros.**

***Palabras claves*— Discurso informal matemático, fenómenos lingüísticos, chequeo automático, expresiones simbólicas.**

Abstract. Informal Mathematical Discourse (IMD) in formal domains, such as mathematics, is characterized by a mixture of natural language and embedded formal expressions. Informal mathematical discourse points in addressing of three major areas of work: (i) The automated proof checking. (ii) The automation of linguistic phenomena occurring in mathematical demonstrations. (iii) The mathematical knowledge management on corpora large of mathematical texts. In this paper, are describes methods, techniques and tendencies on IMD in order to show the applicability of computational linguistic, so that it serves as starting point for development of future projects.

***Keywords*— Informal mathematical discourse, linguistic phenomena, automated proof checking, formal expressions.**

I. INTRODUCCIÓN

El Discurso Informal Matemático (DIM) se caracteriza por la mezcla entre lenguaje natural y expresiones simbólicas que se presentan en el contexto de los libros de textos matemáticos, las publicaciones matemáticas, los diálogos sobre demostraciones matemáticas o la informalidad de una clase de matemáticas. Desde el punto de vista lingüístico, el procesamiento del discurso informal matemático esta direccionado bajo tres aspectos según Wolska [64]: (i) La lingüística, el dominio y el contexto notacional, (ii) la imprecisión del lenguaje informal y lo estructurado de las argumentaciones formales, y (iii) la mezcla de lenguaje simbólico y lenguaje natural. Desde el punto de vista computacional, investigadores como [18],[24],[29],[39],[64],[67] han sido pioneros con sus teorías lingüístico-computacionales y han realizado su mayor esfuerzo en la consolidación de herramientas computacionales que sirvan al procesamiento del lenguaje informal matemático, en la dirección de poder suministrar un conjunto de proposiciones lógicamente formalizadas y verificadas, generando un ambiente de una demostración formal a partir de una informal. La anterior apreciación coincide con lo afirmado por Zinn [67] cuando argumenta que *la semántica de una demostración informal es su correspondiente demostración formal.*

En la actualidad el discurso informal matemático está siendo abordado desde tres grandes áreas de trabajo. Una primera, y quizás la más antigua que hace referencia a la automatización y verificación asistida de la demostración matemática, mediante el uso de los asistentes de demostración matemática, el cual en adelante se llamarán MPAs (*Mathematical Proof Assistants*), en el que se destacan algunos trabajos que buscan el acoplamiento de estos asistentes con el procesamiento del lenguaje natural dando mayor importancia a los fenómenos lingüísticos y a la traducción del lenguaje informal en fórmulas de primer orden para su chequeo y prueba de corrección [8],[12],[13],[27],[28],[30],[33]. Una segunda, que habla de la automatización de los fenómenos lingüísticos que se presentan en los textos matemáticos, publicaciones, demostraciones matemáticas y diálogos en la cual su importancia radica en el procesamiento de los características lingüísticas que son representables usando las técnicas de lenguaje natural tradicional [27],[28],[30],[45],[48],[62],[67]. Y una tercera, que tiene que ver con la Gestión del Conocimiento Matemático, que por su sigla en inglés de aquí en adelante se llamará MKM (*Mathematical Knowledge Management*) y el acceso a grandes corpus de textos matemáticos con el fin primordial de crear un entorno apropiado de herramientas básicas que ayuden al procesamiento de todo tipo de documentos matemáticos en la web [3],[4],[5],[6],[10],[20],[21],[33],[43].

En el artículo se realiza un análisis crítico de la literatura existente sobre el procesamiento del discurso informal matemático, tomando como referencia algunos casos de estudio. El artículo está organizado de la siguiente manera: En la sección 2 se realiza una exposición de los comienzos del DIM y se presenta una revisión de literatura especializada en el procesamiento del DIM mediante la exhibición de los casos de estudio; en la sección 3 se realiza un análisis y discusión de los factores determinantes para el procesamiento del discurso informal matemático; en la sección 4 finalmente, se presentan las conclusiones y el trabajo futuro en el cual estarán orientados todos los esfuerzos del procesamiento del DIM.

II. REVISIÓN DE LA LITERATURA

En la práctica, el procesamiento del DIM es una actividad lenta, consumidora de tiempo y mentalmente muy dura, su inicio se dio por un lado, con el desarrollo de la Matemática Vernacular (MV) (articulación del lenguaje natural con el lenguaje simbólico) y por otro lado, con el desarrollo del Razonamiento Automático (RA). Según Zinn [67] la automatización y procesamiento del DIM requiere de la aplicación de las técnicas del Razonamiento Automático (RA) y el Procesamiento de Lenguaje Natural (PLN). A finales de los años 60's comenzó el auge por el procesamiento del DIM. Trabajos como el de Brobow [9] quien fue uno de los pioneros, lograron definir un subconjunto del inglés para resolver problemas de álgebra mediante el sistema STUDENT, Bobrow [9] tradujo en objetos matemáticos las ecuaciones usando lenguaje natural. Más tarde Cherniak [11] mejoró el trabajo de Bobrow [9] considerablemente, desarrollando el sistema CARPS (*CAlculus Rate Problem Solver*) adicionando mejoras como el refraseo de dos frases para un mismo objeto. Comenzando los 70's De Brujin [15], desarrolló el sistema Automath y definió un lenguaje capaz de representar la expresividad de partes de un texto matemático garantizando la corrección de los contenidos matemáticos, así como la especificación de las reglas gramaticales que lo obedecen. Aunque en los 80's según Zinn [67] no existen trabajos referenciados, las investigaciones sobre el DIM se dieron en el ámbito de los sistemas tutoriales inteligentes. Sistemas como EXCHECK [42] utilizaron gramáticas independientes de contexto para modelar discurso informal matemático y refrasear diferentes expresiones con un mismo significado. Los 90's estuvieron marcados por los trabajos de Simon [55], el cual desarrolló el sistema Nthchecker que transforma una demostración informal a una demostración formal. Paralelamente, otros investigadores como Dowek [16], Groote [24], Luo [39], Ranta [48] realizaron en los 90's su mayor esfuerzo en la consolidación de la MV y suministraron una teoría semántica para la formalización y computarización de las sentencias y pedazos de discurso dentro de un lenguaje informal matemático. También en esta misma dirección y con una MV más reciente, investigadores como Kamareddine [29],[30], Nederpelt [45] y Ranta [49] coadyuvaron al desarrollo del DIM con la definición de estructuras y gramáticas de propósito general para la especificación del lenguaje informal matemático. Por otra parte, en la década del 2000 dos nuevas áreas de investigación han surgido en el procesamiento del DIM: La primera que propende por el uso de las gramáticas de dependencia propuesta por Wolska-Korbayová [62],[63],[64] y la segunda por el uso de la Teoría de Representación del Discurso, en adelante DRT (*Discourse Representation Theory*) propuesta por Kamp-Reyle [31] y desarrollada e implementada por Zinn [67] y por Cramer-Koepke en Naproche [12],[13],[34],[36].

Paralelamente, con el auge de la MV comenzaron las investigaciones sobre el RA y fue en la década de los 60's que Abrahams [1] desarrolló PROOFCHECKER un programa de computador para chequear rigurosas demostraciones en un lenguaje formal y un conjunto restringido de comandos. En la actualidad, el RA mediante el desarrollo de MPAs tales como Nuprl [14], Coq [17], Pvs [52], Hol [26], Mizar [28],[51],[59], Isabelle [46], Automath [15] y ΩMEGA [54] han servido para apoyar a matemáticos en la verificación automática de demostraciones, en la asistencia computarizada de la demostración de teoremas y su corrección. En este sentido, varias investigaciones se han estado llevando a cabo con el fin de integrar los MPAs al DIM. Kamareddine en [29],[30] define una estructura con procesamiento de lenguaje natural de manera automática para todos los MPAs. Benzmüller, Wolska y Horacek en [8] integran ΩMEGA [54] a DIALOG; un sistema tutorial que tiene un conjunto de recursos lingüísticos que incluyen la gramática y el lexicón para el análisis de las entradas en lenguaje natural provenientes de diálogos entre la máquina y el usuario. En esta dirección, se ubica PLATΩ [60] un mediador genérico entre el editor de texto TeXmacs [22],[61] y ΩMEGA [54]. PLATΩ permite a los usuarios escribir sus documentos matemáticos en un lenguaje informal matemático y construye automáticamente la correspondiente representación formal en ΩMEGA [54]. También Humayoun en [27] propone hacer accesible a MathNat a cualquier sistema de chequeo de demostraciones. Finalmente, Kühlwein, Cramer & Koepke en Naproche [12],[13],[34],[36] desarrollaron un lenguaje controlado matemático para el chequeo y verificación lógica de demostraciones. A continuación con una visión más cercana a lo lingüístico y a la estructura discursiva, se muestran algunas de las teorías lingüísticas aplicadas al procesamiento del DIM.

2.1 Teoría de tipos

La Teoría de Tipos (TT) permite formalizar el DIM como un lenguaje en el que se representan los tipos y objetos, ajustados a reglas de inferencia; la adición y la dependencia de tipos sirven para representar las estructuras lógicas. Las proposiciones son representadas como tipos y una prueba de una proposición se define como un término que está en el tipo que representa la proposición. Los teoremas se expresan como proposiciones y probados por un término de algún tipo apropiado; por lo tanto este término puede ser usado como una función para probar otras proposiciones. Hoy en día está la Teoría de Tipos Débil (TTD) de Kamareddine [28],[29],[30] y Nederpelt [45] y la Teoría de Tipos promovidas por De Brujin [15], Lou & Callaghan [39], Groote [24], Ranta [48],[49]. Las diferencias entre la TTD y la TT tienen que ver con que la TTD tiene como objetivo el DIM, mientras que la TT se interesa por el rigor matemático y es más usada por los MPAs. La TTD modela un gran conjunto de nociones matemáticas como: El manejo de constantes, el manejo de ligadores, la diferencia entre proposiciones y conjuntos y las definiciones que son especificadas como ciudadanos de primera clase. Además, la TDD usa un sistema de tipos para representar; la especificación de sustantivos, términos, adjetivos, declaraciones y el contexto de sentencias. Una de las desventajas que presentan la TTD es que no tienen una estructura que permita la representación del discurso. La TT definida por Ranta [48],[49] es una estructura gramatical, de ahora en adelante GF (*Grammar Framework*) que está basada en la teoría de tipos definida por Martin Löf [38], la cual sirve para expresar la semántica del lenguaje natural. Una GF tiene dos componentes: la Sintaxis Abstracta (SA) y la Sintaxis Concreta (SC). La SA define las condiciones abstractas para la formación de los Arboles Sintáctico Abstractos (ASA) y la SC define los objetos lingüísticos asociados a los ASA, suministrando el análisis sintáctico. El proceso de traducción de un ASA a su objeto lingüístico es llamado *linearización.* Dos aplicaciones de la GF se encuentran en MathNath [27] y en el trabajo de Hallgreen [25]**.** Una de las desventajas de la GF es que no sirve para representar el discurso matemático, aunque si puede representar la sintaxis y semántica de cualquier lenguaje informal matemático. Los últimos desarrollos de la teoría de tipos apuntan al trabajo en la web estableciendo para ello el concepto de gramáticas portables, en la búsqueda de formatos gramaticales que sean portables a los diferentes servicios de la web y entre ellos la portabilidad gramatical en el DIM [50].

2.2 Gramáticas de dependencia

El uso de las Gramáticas de Dependencia (GD) en el procesamiento del DIM está relacionado con la utilización de las *Relaciones Tectogramaticales* [53] entre los constituyentes de una expresión matemática a nivel sintáctico-semántico. El trabajo más importante es quizás el desarrollado por Wolska-Korbayová [62] en el cual el significado lingüístico de las expresiones matemáticas se realiza en paralelo con el análisis sintáctico, usando semánticas de dependencia y lógicas híbridas mediante HLDS (*Hybrid logic Dependency Semantics*) y CCG (*Combinatory Categorial Grammar*) [35]. El significado lingüístico representa el significado literal de un enunciado, más que una interpretación en un dominio específico. En este enfoque el significado lingüístico está basado en el análisis profundo mediante la asignación de roles semánticos (funciones gramaticales). Esta asignación de roles se realiza mediante la especificación de las *Relaciones Tectogramaticales* de los dependientes de una cabeza verbo, en la cual estas relaciones de dependencia entre las cabezas y sus dependientes son explícitamente codificados en un lexicón, como relaciones modales. Paralelamente, en este enfoque son tratados fenómenos lingüísticos como: las expresiones imprecisas, la verbalización de algunas acciones en términos y fórmulas, la correferencia en fórmulas y partes de ellas, las expresiones deícticas y la resolución de anáforas. En este sentido, existe en la actualidad un trabajo más refinado de Wolska-Korbayová descrito en [64], cuya arquitectura modular tiene como objetivo activar la parametrización del sistema con respecto al lenguaje natural del texto en cuestión, el dominio matemático del discurso y la notación matemática. La gramática del sistema define un número de categorías sintácticas para las expresiones matemáticas. También se define un componente de interpretación que adiciona significado de dominio específico a los predicados y a las relaciones en un árbol de dependencia. La salida del procesamiento genera un árbol de dependencia semántico con información específica semántica en los niveles de conceptos e interpretaciones del dominio específico. Finalmente, el trabajo de Wolska-Korbayová [62] es quizás el único hasta ahora, en tocar el tema del uso de las relaciones retóricas en la estructura argumentativa de la demostración, relaciones retóricas como *Causa, Condición* y *Resultado-Conclusión* son modeladas en el contexto del DIM, además de otras relaciones encontradas en los TRs como la *Norma-Criterio* y las relaciones sin modificación tales como; *Ubicación, Medios* y *Dirección.* El trabajo futuro de las gramáticas de dependencia y su incidencia al DIM, estará orientado muy probablemente a la definición de estructuras que permitan la desambiguación de expresiones matemáticas y la resolución de anáforas usando corpus como el ARXMLIV [5],[20],[21] y el Zentralblatt Math [66], además del uso de lenguajes abiertos de anotación como el OpenMath [10],[47].

2.3 Uso de las Estructuras de Representación del Discurso DRSs en el DIM.

En la actualidad, la teoría de representación del discurso DRT propuesta por Kamp-Reyle [31] es la más utilizada para el procesamiento del DIM, por su estructura de datos básica y la estructura de representación del discurso. En adelante se hablará de DRS (*Discourse Representation Structure*) que son estructuras que pueden ser adaptadas y extendidas para la representación del DIM. Como lo demuestran los trabajos de Zinn [67] y el de Koepke, Cramer y Kühlwein en el proyecto Naproche [12],[13],[34],[36].

2.3.1 DRSs y verificación de demostraciones informales.

El enfoque definido por Zinn y desarrollado en su tesis doctoral [67], tiene como tarea principal la ejecución un verificador de demostraciones informales, de ahora en adelante VIP (*Veryfing Informal proof)*; el cual inicia con el procesamiento del DIM mediante el análisis sintáctico completo, así como también la construcción de su representación semántica, incorporando a los árboles sintácticos la composicionalidad de los DRSs con λ-términos. Computacionalmente, la construcción de las representaciones semánticas a nivel de sentencias está basada en una versión adaptada de los DRSs, en la cual el resultado del análisis lingüístico es una representación intermedia y subespecificada. Es decir, esta contiene las expresiones referenciales que tienen que ser resueltas en el procesamiento subsiguiente. Todas las sentencias y multisentencias de la demostración son procesadas hasta generar la demostración formal. En general, el VIP puede ser capaz de incorporar más de una representación intermedia dentro del contexto actual de la demostración y también mantener un conjunto de continuaciones de la demostración. La importancia del trabajo de Zinn radica en la representación de las entidades abstractas del discurso matemático mediante la automatización de los referentes abstractos del discurso, usando DRSs; a nivel de sentencia y a nivel discursivo, además de las estructuras de representación de la demostración que en adelante será llamado PRS (*Proof Representation Structure*). Una PRS extiende un DRS en la utilidad y capacidad de los referentes y condiciones del discurso, en la cual las entidades abstractas del discurso son introducidas por un esquema numérico con el fin de imponer un orden en los PRSs. Entre los fenómenos lingüísticos del DIM que cubre la teoría de representación de Zinn están: las construcciones elípticas y anafóricas, las anáforas puente (bridging) y aposicionales, las formas deícticas, la referencia a entidades y el tratamiento de constantes como entidades abstractas. Además de algunos conceptos matemáticos como las proposiciones, constantes, variables, funciones, conectivos, pseudocondicionales y condicionales.

2.3.2 DRSs y Lenguajes Controlados (LC).

Otra de las aplicaciones de los DRSs al DIM se encuentra en Naproche [12],[13],[34],[36], en el cual el objetivo central es desarrollar e implementar un lenguaje controlado para textos matemáticos que puedan ser transformados automáticamente a su fórmula equivalente de primer orden, usando métodos lingüístico-computacionales. En Naproche, los PRSs son una adaptación ampliada de los DRSs que incluyen, como referentes del discurso; los referentes matemáticos, los referentes textuales y los referentes propios del discurso, además de las condiciones de los DRSs. A diferencia de la utilización de λ-DRS y los PRSs en el trabajo de Zinn [67], en Naproche los PRSs utilizan una función de actualización por cada sentencia y se producen nuevos PRSs que sirven de contexto para la generación nuevas funciones de actualización y así sucesivamente, hasta que la lista de sentencias sea revisada. El lenguaje controlado Naproche incluye cierto azúcar sintáctico y cubre la macroestructura de una plantilla de demostración (teorema, axioma, definición, lema y demostración) y la microestructura de las sentencias que conforman la demostración (sustantivos, adjetivos, verbos, sintagmas nominales, sintagmas verbales, etc). Las tendencias del uso de los DRSs en el DIM está enmarcadas: Primero, en el uso de corpus lingüísticos con el fin de informar las interpretaciones de los DRSs en las diferentes lecturas de las sentencias. Segundo, en el uso de las relaciones retóricas en el DIM mediante el estudio de relaciones diferentes a la consecuencia lógica como la elaboración, explicación, narración; partiendo de los DRSs o de sus extensiones.

2.4 Lingüística de corpus

A falta de datos empíricos sobre el uso de lenguaje natural aplicados al discurso informal matemático, se abre paso a la posibilidad de recopilar corpus de diálogos en dominios formales como las matemáticas y corpus de textos matemáticos para su procesamiento en la web [62]. En esta parte del artículo, se presentan tres casos de estudio de los usos de la lingüística de corpus en el procesamiento del DIM: i) Un corpus anotado para diálogos en matemáticas como parte del proyecto DIALOG [8],[63]. ii) El corpus ARXMLIV [5] [20],[21] como transformación del corpus ARXIV [4]. iii) El corpus Zentralblatt MATH [66] como soporte en el estudio de fenómenos lingüísticos en sistemas de gran escala.

2.4.1 Un corpus anotado de diálogos tutoriales.

En el proyecto DIALOG [8],[63], se desarrolló un corpus de diálogos con el fin de estudiar los fenómenos lingüísticos presentes en los diálogos en la enseñanza de la demostración matemática. En DIALOG, se hace uso de la lingüística de corpus con el fin de anotar el significado lingüístico de los enunciados, los movimientos de diálogo y la tutoría. La recolección de los diálogos del corpus fue realizada mediante la herramienta de experimentación *Wizard-of-Oz.* Las anotaciones en el corpus se realizaron con MMAX [44] que soporta la anotación multinivel. En cuanto al lenguaje, los significados lingüísticos de las sentencias están anotados en términos de las relaciones de dependencia semántica en el sentido de la teoría gramatical FGD (*Functional Generative Description*) [53]. Las anotaciones del movimiento del diálogo están basadas en DAMSL [2] y las anotaciones de la tutoría hacen referencia a los consejos del tutor y la categoría de respuestas de los estudiantes. En conclusión, en este corpus están anotados los pasos de demostración y sus justificaciones (métodos de inferencia usados y sus parámetros); premisas y conclusiones; la dirección del razonamiento hacia adelante (forward) o hacia atrás (blackwoard); las relaciones retóricas diferentes a la consecuencia lógica, así como también la representación de las entradas del probador de teoremas.

2.4.2 Corpus para el procesamiento del DIM.

El corpus ARXMLIV [5],[20],[21] es una colección de textos provenientes de ARXIV [4] que contienen discurso matemático científico, con más de medio millón de documentos. El corpus arXMLiv suministra una explícita diferenciación entre lenguaje natural y lenguaje simbólico, facilitando el procesamiento del DIM. En la actualidad, la mayor utilidad de arXMLiv es servir de corpus de experimentación y análisis lingüístico del proyecto LaMaPUn (http://Kwarc.info/projects/lampaun/) en el cual se investiga sobre la desambiguación de fórmulas matemáticas, el procesamiento de fórmulas basados en el contexto y la búsqueda y recuperación de los pasos de demostración en el DIM. Por otra parte, aparece el corpus Zentralblatt MATH [66] que es una base de datos en línea que contiene más de 3 millones de registros procedentes de cerca de 3500 revistas y 1100 publicaciones periódicas desde 1826 hasta la actualidad. Zentralblatt MATH es producido por la editorial FIZ Karlsruhe de Berlin en cooperación con academias europeas e institutos matemáticos. En Zentralblatt MATH se pueden revisar publicaciones matemáticas; datos bibliográficos de publicaciones matemáticas; textos completos de publicaciones matemáticas; y sitios web de instituciones matemáticas. Una de las tendencias en Zentralblatt MATH es la de investigar sobre el tratamiento de la desambiguación de expresiones matemáticas, retomando términos del corpus con la idea de generar un contexto léxico dependiente de este tipo de corpus [23].

2.5 Lenguajes controlados y lenguajes de anotación abiertos

Un lenguaje controlado es un subconjunto del lenguaje natural con sintaxis, semántica y terminología restringidas [65]. Según Humayoun [27] un lenguaje controlado en el DIM permite eliminar la ambigüedad y reducir la complejidad. En MatNath [27] y Naproche [12],[13],[34],[36], los autores en común acuerdo señalan, que la brecha existente entre la informalidad de los libros de texto matemáticos y la formalidad matemática puede ser reducida mediante la inclusión de los lenguajes controlados. Por otro lado, con el fortalecimiento de la MKM han surgido los Lenguajes de Anotación Abierta (LAA) para la representación de formalismos matemáticos en la web, entre ellos los más importantes son: OpenMath [10],[47], MathML (*Mathematical Markup Language*) [6],[41] y OMDoc (*Open Mathematical Documents*) [33].

Los LAA son principalmente lenguajes en la web, que sirven para anotar la información requerida con el fin de recuperar y combinar el conocimiento matemático de manera automática. El aspecto de abierto significa que es extensible con respecto a las nuevas transformaciones de la web y la publicación de textos matemáticos. Lingüísticamente, los LAA son lenguajes estándares extensibles que sirven para representar la semántica de los objetos matemáticos y la pragmática de las sentencias dentro del DIM.

En esta parte del artículo, se presenta el lenguaje controlado MathNat [27], además de la mención que se hizo anteriormente de Naproche. Por el lado de los LAA se encuentran OpenMath, MathML y OMDoc. Finalmente, se mencionan algunos convertidores de formatos [19],[40],[43],[58] en especial de LaTeX [37] a XML (*eXtensible Markup Language*) como soporte al DIM y los LAA. Además, del editor científico TeXtmacs [22],[61].

2.5.1 MathNat.

Es un sistema en el cual se desarrolló un Lenguaje Controlado para Matemáticas (LCM). El lenguaje es un subconjunto de inglés con un vocabulario y gramática restringidas. Para hacer más natural y expresivo el LCM, MathNat soporta algunas características lingüísticas tales como: anáforas, correferencias, refraseo de una sentencia en múltiples maneras y manejo distributivo y colectivo de lecturas correctas. Otra de las actividades de MathNat es que traduce el LCM a un sistema independiente de lenguaje formal MathsAbs con la idea de hacer MathNat accesible a sistemas de chequeo de demostraciones (*proof checking*). Actualmente se está traduciendo MathAbs a su equivalente en fórmulas de primer orden para la verificación y corrección. El texto de LCM es una colección de axiomas, definiciones, proposiciones teoremas, lemas y demostraciones estructuradas; especificadas por palabras claves como Axioma, Definición, Teorema, Demostración, etc. Estas palabras claves son listas de sentencias que obedecen a diferentes gramáticas de tipo GF [49] (una para axiomas, una para definiciones, etc) y a su vez esas gramáticas no son independientes pero comparten reglas comunes.

2.5.2 Los LAA en el DIM.

Los LAA surgen como herramientas de anotación semántica dentro del DIM. Entre los más importantes LAA está OMDoc [33], que soporta el contenido matemático en tres niveles sobre la anotación semántica y pragmática del discurso matemático. En el *nivel objeto de fórmulas,* OMDoc usa los estándares establecidos de Openmath [10],[47] y MathML [6],[41], ambos suministran anotación de contenido para la estructura de fórmulas matemáticas y anotación de contenido en la forma de referencias URI (*Uniform Resource Identifier*) para la representación de símbolos. En *el nivel de sentencia;* de definiciones, teoremas, demostraciones, ejemplos y relaciones entre ellos, OMDoc suministra una infraestructura de anotaciones originales para hacer explícitas las sentencias, teniendo en cuenta el contenido y los aspectos de anotación de contexto. En el *nivel teórico,* OMDoc suministra una anotación original para conjuntos de sentencias agrupadas por teorías, especificando relaciones entre teorías por morfismos. Una importante distinción en OMDoc que marca la diferencia entre la matemáticas formal e informal, son la *anotación de contenido* y la *anotación semántica (sistema formal);* este último formato se define como un sistema de representación que tiene una forma de especificar cuando una fórmula es una consecuencia de otra. La *anotación de contenido* en cambio, no comprende una particular relación de consecuencia, y se concentra en la prestación de servicios basados en la estructura de marcado del contenido y el contexto. Una de los usos de OMDoc se presenta en [7], un experimento que utiliza el sistema Mizar [51],[59] para transformar la librería de Mizar a un formato OMDoc. El segundo LAA es MathML [6],[41] que es una aplicación XML para describir la notación matemática y capturar su estructura y contenido. El objetivo de MathML es permitir que las matemáticas sean funcionales y procesadas en la web. MathML ofrece dos sublenguales; *pre*sentación y *contenido* MathML. El primero anota la apariencia visual de las fórmulas representándolas como un árbol de primitivas. El segundo anota la estructura funcional del contenido de las fórmulas matemáticas.

El tercer LAA es OpenMath [10],[47], que es un estándar para la representación y comunicación entre objetos matemáticos. En OpenMath el significado de un objeto puede ser codificado más que su representación visual. El formato OpenMath no es principalmente una aplicación XML, por el contrario OpenMath define un *modelo de objeto abstracto* para objetos matemáticos y especifica una codificación binaria XML. Esta codificación binaria permite dos cosas; optimizar el tamaño de la codificación y más importante aún, optimizar el tiempo del análisis sintáctico para grandes objetos de OpenMath. El manejo de los símbolos que se utilizan para representar la gran cantidad de constantes del dominio matemático es tal vez la gran diferencia entre OpenMath y contenido-MathML. En lugar de suministrar todos los elementos para los conceptos k-14 (conjunto de glifos de uso general para símbolos matemáticos) como lo hace MathML, el estándar de OpenMath adiciona un mecanismo de extensión para conceptos matemáticos, *los diccionarios de contenido.* En resumen, se puede decir que MathML es el formato de presentación preferido por los objetos de OpenMath y los *diccionarios de contenido* de OpenMath son el lenguaje de especificación principal para la semántica de MathML.

2.5.3 Convertidores, editores y motores de búsqueda.

Para muchos, LaTeX [32] es el formato preferido para la creación de documentos que involucran contenido matemático y donde la calidad simbólica y tipográfica son importantes. Por otro lado, XML es un lenguaje de marcado extensible que nos permite la rehusabilidad con diversos propósitos en la web. Una de las grandes alternativas en la discriminación del contenido matemático con referencia al contenido textual de documentos, es la conversión de LaTeX a XML. LaTeXML [43], es un módulo en Perl que analiza sintácticamente un documento de LaTeX y emite un documento de salida en XML. LaTeXML consiste de un analizador de TeX [56], un emisor XML y un postprocesador pipeline. Actualmente, LaTeXML es capaz de generar la *presentación* y *contenido* de MathML, así como también una representación matemática de OpenMath. A nivel de grandes corpus, la tendencia del uso de LaTeXML está orientada a la conversión de grandes corpus como en el caso de la conversión de arXiv [4] a arXMLiv [5] bajo el contexto del discurso matemático. Otros convertidores son Hermes [3], GELLMU [19], LXir [40], TeX4ht [57], Tralics [58]. Otra de las herramientas usadas en el procesamiento del DIM es TeXtmacs [22],[61] que facilita la definición de macros como como lo hace LaTeX, permitiendo la definición de plugins para la automatización, disposición e integración de documentos. En la actualidad, el sistema Naproche está diseñado como un plugins de TeXmacs en el cual un módulo de entrada en Naproche crea un archivo XML a partir de la entrada en TeXmacs [34],[61]. Otra de la aplicaciones de TeXmacs se referencia en PLATΩ [60], en el que también fue desarrollado un plugins que establece una conexión entre TeXmacs y PLATΩ.

III. ANÁLISIS Y DISCUSIÓN

Muchas veces las ideas matemáticas están expresadas y escritas de manera informal, guardando un carácter lógico-deductivo de igual semántica que en el lenguaje formal. No existe un común acuerdo o una sintaxis prescrita sobre cómo se deben expresar las ideas alrededor del DIM. Para el análisis y discusión del tema, se adoptará el nombre genérico de lenguaje informal matemático como el lenguaje propio del DIM, el cual ha existido siempre con el fin de expresar el contenido matemático y comunicarlo a su entorno académico correspondiente. Desde los 60's con los trabajos de la MV y el RA hasta nuestros días, se habla de un lenguaje informal matemático en el cual se expresa el mundo matemático y que en algunas veces se convierte en lenguaje controlado cuando se restringe la gramática del lenguaje, aunque no se hace explícita la característica de control. Por una parte, los lenguajes informales matemáticos se preocupan por la cantidad y calidad de los fenómenos lingüísticos que alcanzan a cubrir, sin embargo, a veces en el procesamiento de estos lenguajes no se tienen en cuenta problemas como: la ambigüedad, refraseo, metonimia y el fenómeno anafórico en todas sus dimensiones. Por otra parte, los lenguajes controlados en el DIM permiten controlar el conjunto de instancias del lenguaje simbólico dentro del lenguaje natural, restringiendo la gramática y soportando las características lingüísticas de refraseo de sentencias, referenciación anafórica, eliminación de ambigüedad de manera anticipada y manejo de lecturas colectivas. De lo anterior se puede deducir que ambos tipos de lenguajes propenden por la expresividad del lenguaje mismo, bajo la perspectiva del procesamiento complejo de los fenómenos lingüísticos. En este contexto se mueven todas las teorías lingüístico-computacionales antes mencionadas. Por lo tanto, la discusión en el procesamiento del DIM se debe orientar a la expresividad que puede ser generada por la automatización de los lenguajes informales matemáticos y dar respuesta a preguntas como ¿Qué tan expresiva puede ser la automatización del lenguaje? o ¿Cuál es la relación entre la expresividad y el cubrimiento de los fenómenos lingüísticos por parte de las teorías lingüístico- computacionales? En este sentido, varios investigadores [27],[45],[62], aseguran que la expresividad de los lenguajes informales matemáticos es comparable con las estructuras computacionales que han sido definidas dentro del procesamiento del DIM, sin embargo, Nederpelt [45] va más allá y expresa que estos lenguajes son también apropiados para expresar todo tipo de modelos mentales matemáticos que van desde las entidades matemáticas y las relaciones entre ellos, al razonamiento matemático y teorías, con la capacidad de establecer diferentes categorías matemático-lingüísticas. La expresividad sobre la automatización de los lenguajes informales matemáticos se garantiza siempre y cuando el conjunto de fenómenos lingüísticos del DIM sea cubierto mediante el uso de las teorías lingüístico-computacionales. También es claro que las teorías provenientes del procesamiento de lenguaje natural con sus adaptaciones han ayudado a la expresividad de los lenguajes informales matemáticos, teorías como los DRSs y las gramáticas de dependencia han servido de base y estructura para el modelamiento de fenómenos lingüísticos complejos. El hecho a revisar en esta discusión tiene que ver con la relación entre la expresividad y los fenómenos lingüísticos que se pueden cubrir en el DIM, para ello se establece un cuadro comparativo sobre que fenómenos lingüísticos son cubiertos por cada una de las teorías lingüísticas propuestas en el apartado anterior con sus respectivas referencias.

Tabla 1. Cubrimiento de los fenómenos lingüísticos y las teorías computacionales

		TEORIAS LINGÜÍSTICO-COMPUTACIONALES				
		DRS	GD	LC	LAA	TT y MV
FENOMENOS LINGÜÍSTICOS	*Ambigüedad	[67]		[12],[27]	[33]	
	*Anáforas	[34],[67]	[64]	[27]		[27]
	Correferencias	[67]	[62]	[12],[27],[64]	[6],[33][41]	[6],[30]
	Metonimia		[62],[64]			
	Verbalización: Sintagmas, adjetivos, adverbios,	[67]	[62],[64]	[12],[27],[30]	[6],[33],[41]	[1],[9],[11] [15],[16],[18] [24],[30],[39] [48],[55]
	Refraseo de sentencias	[67]	[62],[64]	[27]		[9],[11],[42]
	*Relaciones retóricas		[64]			[30]
	Uso de corpus		[63],[64]		[4],[5], [10],[33] [43],[47], [66]	

En la tabla 1 los fenómenos lingüísticos que están con asterisco son los más complejos en cuanto a su implementación y formalización, en la tabla también aparecen las referencias por teorías y los fenómenos que cubren. La interpretación que se puede hacer es tan relativa como la medición de la expresividad en el procesamiento del DIM, por ejemplo; los DRSs aplicados al lenguaje informal matemático de los textos matemáticos, facilitan el tratamiento de las referencias anafóricas en todas sus dimensiones (Aposicional, puente (*Bridging*), formas deícticas). Los DRSs por su estructura y por lo explícito de los referentes del discurso facilitan la representación de las anáforas en todas sus formas, caso contrario sucede con las gramáticas de dependencia, los lenguajes controlados y la teoría de tipos en la cual el tratamiento de la anáfora es solamente pronominal y se debe recurrir a algoritmos mucho más complejos. Por el lado de la ambigüedad, los lenguajes controlados no la tratan de manera directa pero si la evitan, basándose en el control y restricciones de las reglas gramaticales, permitiendo delimitar el concepto de constantes, variables y sentencias ambiguas. Los otros fenómenos lingüísticos como la metonimia, la verbalización y correferencias son mejor tratados por las gramáticas de dependencia en el contexto de los diálogos matemáticos, aunque la articulación de los DRSs y el cálculo λ bajo el principio de composicionalidad es una herramienta capaz de enfrentar la mayoría de estos fenómenos. Con respecto a las relaciones retóricas, estas son modeladas en la teoría de tipos mediante el uso de la teoría de grafos [30] y mediante el uso de las relaciones tectogramaticales [62]. Otro aspecto relevante que sirve de soporte a la expresividad de los lenguajes informales matemáticos es el conjunto de características matemáticas que pueden ser representadas por las teorías lingüístico-computacionales en cuestión, en su mayoría estas teorías representan casi todos las características del lenguaje formal como son la representación de constantes, variables, funciones, predicados y condicionales. El alcance de cuantificación de las variables es quizás el más complejo, el cual sigue siendo un tema de investigación pero que ha sido tratado con relativo éxito. Por ejemplo en la teoría de tipos el alcance es modelado para cada línea del texto de demostración de manera diferente al alcance de las variables que están definidas en las estructuras macro de la demostración (axiomas, definiciones, lemas, etc). En los DRSs el alcance de cuantificación es representado mediante la composicionalidad de los componentes de las sentencias y la aplicación del cálculo λ. Un último tema que toma relevancia es la lingüística de corpus aplicado al DIM que sólo es notable en el trabajo de Wolska [63],[64] con el fin de anotar algunas relaciones retóricas y en los trabajos de transformación de grandes corpus como el ARXIV [4] y el uso del corpus Zentralblatt MATH [66] en el procesamiento de la ambigüedad de expresiones en textos matemáticos.

IV. CONCLUSIONES Y TRABAJO FUTURO

Una primera conclusión que se puede extraer sobre la expresividad de los lenguajes informales matemáticos, se da en el entorno del cubrimiento de los fenómenos lingüísticos el cual dependen del tipo de DIM (diálogos matemáticos, textos matemáticos, corpus de demostraciones, etc). Por lo tanto, la prioridad de estos fenómenos obedece a la estructura misma del lenguaje informal matemático. Otra de las conclusiones importantes tiene que ver con el uso obligatorio de las estructuras articuladas para la representación del DIM. Por ejemplo, en el contexto de los diálogos matemáticos se articulan las relaciones tectogramaticales con las gramáticas categoriales y las lógicas modales, dando como resultado el tratamiento de metonimia, correferencias, refraseo y relaciones retóricas, (Tabla 1: GD vs fenómenos lingüísticos), descuidando un poco las referencias anafóricas. En el contexto de la automatización de los textos matemáticos se articulan los DRSs con el cálculo λ, con el fin de representar, las relaciones anafóricas en su totalidad, alguna ambigüedad, correferencias y el refraseo (tabla1: DRS vs fenómenos lingüísticos). Otro ejemplo de estructura articulada, que viene desde la década de los 80's es la articulación de la MV y la TT, para el procesamiento de fenómenos lingüísticos de menor complejidad como: verbalizaciones, refraseo, correferencias y algunas anáforas (tabla1: TT y MV vs fenómenos lingüísticos). Finalmente, hoy en día la comparabilidad de los lenguajes informales matemáticos con las estructuras lingüístico-computacionales existentes, acortan la brecha existente en temas propios del DIM como la informalidad, la ambigüedad, la incompletitud, la pobre organización de textos, las demostraciones y diálogos matemáticos y la automatización de algunos procesos propios del razonamiento humano.

Varias son las líneas de trabajo que requieren de una especial atención con relación al procesamiento del DIM y que pueden generar proyectos futuros realizables dentro del marco de las teorías lingüístico-computacionales. Por ejemplo, el desarrollo y anotación de corpus para diferentes dominios de las matemáticas con el fin de estudiar y analizar los fenómenos lingüísticos del DIM en los niveles sintáctico, semántico y discursivo. Particularmente, un corpus de textos matemáticos podría servir para investigar la semántica y la pragmática de palabras claves o conectivos (por lo tanto, entonces, o , y) [67]. El análisis del corpus además se complementaría con un estudio cuantitativo de la frecuencia de uso de estas palabras claves (*por ejemplo ver el porcentaje de utilización del conectivo lógico entonces o el conectivo por lo tanto*). También en la lingüística de corpus falta la creación de un gran corpus anotado en español, que sirva como corpus de experimentación para los casos de uso de demostraciones, textos, diálogos pedagógicos en el que se puedan utilizar la recuperación de información, el aprendizaje de máquinas y la semántica superficial. En este sentido el trabajo futuro también estará orientado en conseguir que las herramientas para la anotación sintáctica, semántica y discursiva estén más enfocadas a contrarrestar muchos de los problemas del procesamiento del DIM. Otro de los trabajos futuros y apasionantes tiene que ver con el estudio de las relaciones retóricas en el DIM. En el momento no existe una teoría formal propia para el manejo de las relaciones retóricas diferentes a las relaciones de causalidad, conclusión y condicionalidad expuestas en [62]. Un estudio de las relaciones retóricas en el DIM desde la perspectiva de las teorías del procesamiento de lenguaje natural sería de gran importancia en la consecución de una representación mucho más estructurada del discurso. Sin embargo, aún no es claro si las relaciones retóricas (explicación, elaboración, narración, precedencia) pueden aparecer en el discurso de los textos matemáticos. Esto no es contrario a la realización de un corpus de estudio de demostraciones matemáticas en la cual se puedan anotar discursivamente las relaciones retóricas. Un último trabajo futuro estaría enmarcado en la investigación sobre los LAA y el DIM, sería interesante orientar la investigación sobre la formalización pragmática y discursiva del DIM mediante la articulación de lenguajes como OMDoc y los DRTs usando la lingüística de corpus para dominios matemáticos pequeños.

Finalmente, el uso de nuevos formalismos gramaticales y su articulación a las teorías de representación discursiva; el uso de los lenguajes controlados como una alternativa para generar más expresividad; el desarrollo de eficientes herramientas para el procesamiento de grandes volúmenes de textos matemáticos sobre corpus y la estructuración del discurso mediante el uso de las teorías tradicionales del procesamiento de lenguaje natural junto con los adelantos de la lingüística computacional; seguirán siendo los grandes desafíos en el procesamiento del DIM.

REFERENCIAS

[1] Abrahams P.W., 1963. Machine Verification of Mathematical proofs. MIT. PhD Thesis.
[2] Allen J.; Core M., 1997. Draft of DAMSL: http://www.cs.rochester.

edu/research/speech/damsl/RevisedManual/
[3] Anghelache R., 2007. Hermes: A semantic xml+mathml+unicode epublishing/ selfarchiving tool for latex authored scientific articles. http://hermes.roua.org/
[4] ArXiv. 2011. Cornell University Library. http://arxiv.org/
[5] ArXMLiv. 2011. http://arxmliv.kwarc.info/
[6] MathML. 2003. Mathematical Markup Language (MathML) version 2.0. W3C. http://www.w3.org/TR/MathML2/
[7] Bancerek G.; Kohlhase M., 2007. Towards a Mizar mathematical library in OMDoc format.
Studies in logic, Grammar and rethoric 10 (23).
[8] Benzmüller C.; Fiedler A.; Gabsdil M.; Horacek H.; Kruijff-Korbayová; Pinkal M.; Siekmann J.; Tsovaltzi D.; B. Q. Vo; y Wolska M., 2003. Tutorial dialogs on mathematical proofs. En: IJCAI Workshop on Knowledge Representation and Automated Reasoning for E-Learning Systems.
[9] Bobrow D.G., 1964. Natural language input for a computer problem solving system. PhD thesis. MIT.
[10] OpenMath. 2006. The Open Math standard, version 2.0. The Open Math Society.
http://www.openmath.org/index.html.
[11] Charniak E., 1969. Computer solution of calculus word problems. First International Joint Conference on Artificial Intelligence, pp. 303-316.
[12] Cramer M.; Koepke P.; Kühlwein P,; Schröder B.; Veldman J., 2009. The Naproche Project Controlled Natural Language Proof Checking of Mathematical Texts. CNL´09 Proceedings of the 2009 conference on Controlled natural language. Springer Verlag.
[13] Cramer M.; Koepke P.; Kühlwein D.; Schröder S., 2010. Premise selection in the naproche system. J. Giesl and R. Hähnle (Eds.): IJCAR 2010, LNAI 6173, Springer Verlag Berlin Heidelberg, pp. 434–440.
[14] Constable R.L., 1986. Implementing mathematics with the nuprl proof development system. Prentice Hall.
[15] De Brujin N.G., 1970. The mathematical Language Automath, its Usage, and Some of its Extensions. In symposium on Automatic Demostration, M. Laudet et al. (eds), volume 125 of Lectures Notes in Mathematics, Springer Verlag, pp. 29-61.
[16] Dowek G., 1990. Naming and Scoping a Mathematical Vernacular. Technical Report 1283 INRIA. http://hal.inria.fr/inria-00075276/en/
[17] Dowek G.; Felty A.; Herbelin H.; Huet G.; Pauli Mohring C.; Werner B., 1991. The Coq proof assistant, version 5.6 user's guide. Technical report, INRIA.
[18] Fox C., 2001. Vernacular Mathematical: Discourse Representation and Arbitrary Objects. London University.
[19] GELLMU. www.albany.edu/~hammond/gellmu.
[20] Ginev D., 2009. Semantic Purification of arXMLiv Mathematical Fragments. Jacobs University Bremen, Germany.
[21] Ginev D.; Jucovschi C.; Anca S.; Grigore M.; David C.; Kohlhase M., 2009. An Architecture for Linguistic and Semantic Analysis on the ARXMLIV Corpus. In Applications of Semantic Technologies (AST) Workshop at Informatik.
[22] GNU TeXmacs., 2003. http://www.texmacs.org/
[23] Grigore M.; Wolska M.; Kohlhase M., 2009. Towards context-based disambiguation of mathematical expressions. En: Proceedings of the Asian Symposium on Computer Mathematics.
[24] Groote J.F.,1992. Towards a Formal Mathematical Vernacular. University of Utrecht.
[25] Hallgren T.; Ranta A., 2000. An extensible proof text editor. Springer. 1955.
[26] Hol. 2003. http://hol.sourceforge.net
[27] Humayoun M.; Rafalli C., 2008. MathNat- Mathematical Text in a Controlled Natural Language. Special issue: Natural Language Processing and its Applications. Journal on Research in Computing Science. Volume 46. ISSN:1870-4069.
[28] Kamareddine F.; Maarek M.; Retel K.; Wells J.B., 2007. Gradual Computerisation/ Formalisation of Mathematical Texts into Mizar. Studies in logic, grammar and rhetoric 10 (23).
[29] Kamareddine F.; Wells B.J., 2007. Restoring Natural Language as a Computerised Mathematical Input Method. En: proceedings Calculemus ´07/MKM ´07. Springer Verlag.
[30] Kamareddine F.; Wells J.B.; Zengler C., 2008. Computerising Mathematical Text with MathLang. Electronic Notes in Theorical Computer science (ENTCS). Volume 205. Elsevier Science Publishers.
[31] Kamp H.; Reyle H., 1993. From Discourse to Logic. Kluwer Academic Publishers.
[32] Knuth D. E., 1984. The TEXbook. Addison Wesley.
[33] Kohlhase M., 2010. An Open Markup Format for Mathematical Documents. OMDoc [Version 1.2]. Creative Commons Share-Alike license. https://trac.omdoc.org/OMDoc
[34] Kolev N.M., 2008. Generating Proof Representing Structures in the Project Naproche. Magister Artium M.A. http://naproche.net/downloads/2008/2008%20kolev%2001.pdf
[35] Krujiff G.J.; Krujiff-Korbayová I., 2001. A Hybrid Logis Formalization of Information Structure Sensitive Discourse Interpretation. En: Proceedings 4th International Conference on Text, Speech and Dialogue (TSD'2001). Springer Verlag.
[36] Kühlwein D., 2008. A calculus for Proof Representation Structures. http://www.naproche.net/inc/downloads.php
[37] Lamport L., 1994. LATEX: A Document Preparation System, 2/e. Addison Wesley.
[38] Löf M., 1984. Intuitionistic Type Theory. Napoli, Bibliopolis.
[39] Luo Z.; Callaghan P., 1997. Mathematical Vernacular and conceptual well-formedness in mathematical language. En: proceedings LACL ´97 Selected papers from the second International Conference on Logical Aspects of Computational Linguistics. Springer Verlag.
[40] LXir., 2007. www.lxir-latex.org
[41] MathML., 2010. http://www.w3.org/TR/MathML3/
[42] McDonald J., 1982. The EXCHECK CAI System. En: Suppes, P. (Ed.) University-level Computer-assisted Instruction at Stanford: 1968-1980. Institute for Mathematical Studies in the Social Sciences, Stanford University, Stanford, California.
[43] Miller B., 2007. LaTeXML: A LATEX to XML Converter. Web Manual en http://dlmf.nist.gov/LaTeXML/
[44] Müller C.; Strube M., 2003. Multi-level annotation in MMAX. En: Proceedings of the 4th SIGdial Workshop on Discourse and Dialogue, Sapporo. Japón.
[45] Nederpelt R.P., 2002. Weak Type Theory: A Formal Language for Mathematics, (Computer Science Report, No. 02-05). Eindhoven: Technische Universiteit Eindhoven, pp. 74.
[46] Nipkow T.; Paulson L.C.; Wenzel M., 2002. Isabelle/HOL | A Proof Assistant for Higher-Order Logic, volume 2283 of LNCS. Springer Verlag.
[47] OpenMath., 2006. http://www.openmath.org/
[48] Ranta A., 1994. Type Theory and the Informal Language of Mathematics. Lecture Notes in Computer Science, Volume 806, pp. 352-365.

[49] Ranta A., 2003. Grammatical Framework: A Type Theoretical Grammar Formalism. *Journal of Functional Programming*, 14(2), pp. 145-189.
[50] Ranta, A.; Bringert B.; Angelow K., 2009. Grammatical Framework Web Service. En: European Chapter of the Assocoation for Computational Linguistic. (EACL). The Association for Computational Linguistics (ACL).
[51] Rudnicki P., 1992. An Overview of the MIZAR Project. En: Proceedings of the 1992 Workshop on Types for Proofs and Programs, Bastad.
[52] Rushby J., 2002. PVS bibliography. Technical report, Computer Science Laboratory, SRI International, Menlo Park, California, U.S.A, Constantly updated; available at http://www.csl.sri.com/users/rushby/pvs-bib.html
[53] Sgall, P.; Hajicová E.; Panevová J., 1986. The Meaning of the Sentence in its Semantic and Pragmatic Aspects. Kluwer Academic Publishers Group.
[54] Siekman J.; Benzmüller C.; Autexier, S., 2006. Computer Supported Mathematics with ΩMEGA. Journal of Applied Logic 4, pp. 533-559.
[55] Simon, D.L., 1990. Checking Number Theory Proofs in Natural Language. PhD thesis. UT Austin.
[56] TeX., 2011. www.tug.org
[57] TeX4ht., 2011. http://www.tug.org/tex4ht/
[58] Tralics., 2011. http://www-sop.inria.fr/miaou/tralics/
[59] Trybulec, A., 1978. The MIZAR-QC/6000. Logic Information Language. ALLC, 6, pp. 136-140.
[60] Wagner M.; Auxtexier S.; Benzmüller C., 2006. PLATΩ: A Mediator between Text-Editors and Proof Assistance Systems. En: Proceedings of the 7th Workshop on User Interfaces for Theorem Provers (UITP 2006), volume 174(2) of ENTCS, pp. 87-107. Elsevier.
[61] Wagner M.; Lesourd H., 2008. Using TEXMACS in Math Education: An exploratory Study. Workshop on Mathematical User Interfaces (MathUI'08).
[62] Wolska M.; Krujiff-Korbayová I., 2004. Analysis of Mixed Natural and Symbolic Language Input in Mathematical Dialogs. En: ACL '04 Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics.
[63] Wolska M.; Krujiff-Korbayová I.; Bao Quoc Vo; Tsovaltzi D.; Karagjosova E.; Horacek H.; Fiedler A.; Benzmüller C., 2004. An Annotated Corpus of Tutorial Dialogs on Mathematical Theorem Proving. En: COLING '08 Proceedings of the 22nd International Conference on Computational Linguistics.
[64] Wolska, M., 2008. A Language Engineering Architecture for Processing Informal Mathematical Discourse. Towards Digital Mathematics Library. Birmingham, United Kingdom, pp. 131-136.
[65] Zapata C.; Rosero R., 2008. Revisión crítica de la literatura especializada en lenguajes controlados. Revista Avances en Sistemas e Informática. Vol. 5 No. 3. ISSN 16577663.
[66] Zentralblatt Math., 2010. Project home page at http://www.zentralblatt-math.org/zmath/
[67] Zinn C., 2004. Understanding Informal Mathematical Discourse. PhD thesis, Universität ErlangenNürnberg Institut für Informatik.