

ALVARO GARCÍA

La terminología

en los países

andinos



Director de la publicación: Bernardino Osio

Jefe de Redacción: Daniel Prado

Coordinación de este número: María Eugenia Franceschi, Alejandra Segui

Colaboradores: Bertha Nelly Cardona R. (bcardona@carios.udea.edu.co), Hilda Y. Contreras (hyelitza@ula.ve), Jacinto A. Dávila Q. (jacinto@ula.ve), Alicia Fedor de Diego (alicia_fedor@yahoo.com), María Eugenia Franceschi (e.franceschi@unilat.org), Pilar García Donoso (pilargd@accessinter.net), Iván Guzmán de Rojas (igr@atamiri.cc), Carlos Leáñez Aristimuño (ulatven@cantv.net), Rosa Luna (rosalunacl@terra.com.pe), Melva J. Márquez R. (melva@ing.ula.ve), Tamara Montero (monerot@pdvsa.com), María Teresa Múnera Torres (mmuner@hotmail.com), Plutarco Naranjo (plutarco@lenguaje.com), Gabriel Angel Quiróz Herrera (malergia@tutopia.com), Emma Rodríguez Camacho (erodrigu@univalle.edu.co), Ivana Suito Buselli (ivasuito@terra.com.pe).

Revisión general: María Eugenia Franceschi

Resúmenes: Inés Bravo, Yves Goubert

Traductores y revisores: Inés Bravo, Renée Eve Levié, Céline Normier, Fernando dos Santos, Claudia Valeiro

Concepción gráfica: Alejandra Segui

Diseño de tapa: Bettina Pell

Depósito legal: octubre de 2002

ISSN: 1683-7673

Impreso en: Policrom S.A. - Tánger, 25 - 08018 Barcelona - ESPAÑA

❖ ÍNDICE

❖ PRESENTACIÓN	1
I ❖ LA TERMINOLOGÍA EN LOS PAÍSES ANDINOS	
La terminología en Colombia: panorama general y algunas pautas históricas	3
Lugar que ocupa la terminología en la formación de traductores en el Perú	7
Sobre las actividades terminológicas en Venezuela	9
La terminología y la ingeniería lingüística en la Universidad de Los Andes	16
La Escuela Interamericana de Bibliotecología en la historia de la terminología en Colombia	23
Colterm, Red Colombiana de Terminología	26
Aleterm, Red de Terminología América Latina-España	29
Red Iberoamericana de Terminología	30
VenTerm, Asociación Venezolana de Terminología	31
La terminología en la formación de un traductor especializado	33
La terminología en el proyecto de maestría en información y documentación de la Escuela Interamericana de Bibliotecología	38
Traducción y terminología en la Pontificia Universidad Católica del Ecuador	43
Diplomado en terminología y comunicación especializada (Universidad del Valle)	46
II ❖ LAS INDUSTRIAS DE LA LENGUA EN LOS PAÍSES ANDINOS	
La Torre de Babel en Internet	49
Unifeterm, proyecto de banco de datos terminológico	50
La gestión terminológica en la gerencia de conocimiento de PDVSA Intevep	54
El motor morfológico de Signum y sus aplicaciones	58
Introducción a la lingüística computacional con una perspectiva interdisciplinaria	65
III ❖ REFERENCIAS	72

❖ ÍNDICE

❖ PRESENTACIÓN	1
I ❖ LA TERMINOLOGÍA EN LOS PAÍSES ANDINOS	
La terminología en Colombia: panorama general y algunas pautas históricas	3
Lugar que ocupa la terminología en la formación de traductores en el Perú	7
Sobre las actividades terminológicas en Venezuela	9
La terminología y la ingeniería lingüística en la Universidad de Los Andes	16
La Escuela Interamericana de Bibliotecología en la historia de la terminología en Colombia	23
Colterm, Red Colombiana de Terminología	26
Aleterm, Red de Terminología América Latina-España	29
Red Iberoamericana de Terminología	30
VenTerm, Asociación Venezolana de Terminología	31
La terminología en la formación de un traductor especializado	33
La terminología en el proyecto de maestría en información y documentación de la Escuela Interamericana de Bibliotecología	38
Traducción y terminología en la Pontificia Universidad Católica del Ecuador	43
Diplomado en terminología y comunicación especializada (Universidad del Valle)	46
II ❖ LAS INDUSTRIAS DE LA LENGUA EN LOS PAÍSES ANDINOS	
La Torre de Babel en Internet	49
Unifeterm, proyecto de banco de datos terminológico	50
La gestión terminológica en la gerencia de conocimiento de PDVSA Intevep	54
El motor morfológico de Signum y sus aplicaciones	58
Introducción a la lingüística computacional con una perspectiva interdisciplinaria	65
III ❖ REFERENCIAS	72

ANEXO 3

REGISTRO DE TRABAJOS DE GRADO UNIMET

Título	Autor	Tutor	Fecha de discusión
Análisis terminológico para fines didácticos de los conceptos inherentes al área de filosofía educacional, pedagogía y psicología educacional, dirigido a los estudiantes de Idiomas Modernos de la Universidad Metropolitana.	Rosa Sánchez Corona y Rita Sano Cappuccio	Alicia Fedor de Diego	1988
Glosario básico español-inglés de los términos legales utilizados en el área de dumping en Venezuela	Claudia Capitano Alvarez Mercedes Alonso	Tatiana Lugo	2000
Glosario terminológico (español-inglés) de automatización de equipos industriales	Yanira Briceño Morales	Tatiana Lugo	2001
El Caballo Purasangre de Carreras (PSC) en Venezuela - Prototipo de un glosario terminológico	María Esperanza Miglietti	Alicia Fedor de Diego	2002

La terminología y la ingeniería lingüística en la Universidad de Los Andes: historia y situación actual

Cet article offre un panorama général sur le travail mené au cours de ces 30 dernières années à l'Université de Los Andes en matière d'enseignement et de recherche dans le domaine de la terminologie. Il décrit également les derniers progrès ayant permis de faire face à la nécessité de conjuguer tous les efforts académiques dans un contexte de travail interdisciplinaire, notamment avec l'ingénierie linguistique, où la terminologie et la documentation occupent une place privilégiée. L'article présente, enfin, une liste des produits terminologiques et lexicographiques publiés par des membres de l'Université de Los Andes.

Este artigo apresenta um panorama geral do trabalho realizado nos últimos 30 anos na Universidade de los Andes em matéria de ensino e pesquisa na área da terminologia. Ele também descreve os últimos progressos que possibilitaram conjugar todos os esforços acadêmicos num contexto de trabalho interdisciplinar, sobretudo com a engenharia lingüística, onde a terminologia e a documentação têm um lugar privilegiado. Por último, o artigo apresenta uma lista dos produtos terminológicos e lexicográficos publicados pelos membros da Universidade de los Andes.

Introducción

El presente artículo persigue dos objetivos fundamentales: en primer lugar, ofrecer una visión panorámica del trabajo, tanto en docencia como en investigación, que se ha realizado en la Universidad de Los Andes dentro del área de la terminología en los últimos 30 años, especial-

mente desde que se creó el Servicio de Traducción en 1974. En segundo lugar, el artículo da a conocer los acuerdos, proyectos alcanzados y en curso, donde áreas como la terminología, la documentación y la lingüística computacional ocupan un lugar preponderante. En este apartado se destaca la necesidad de vincular áreas estrechamente relacionadas como la informática, la lógica, el análisis del

discurso, la lingüística cognitiva, la lingüística textual y la documentación. De esta manera, se empieza a abrir un abanico temático de investigación y docencia que puede obtener excelentes resultados tanto para la comunidad universitaria como para el resto de la sociedad.

El artículo está dividido en los siguientes apartados: el primero ofrece una breve descripción de la Universidad de Los Andes, su estructura física, académica y humana, y un recorrido por la historia de la institución. El segundo trata del estudio de la terminología en la Universidad durante los últimos 25 años, tanto en docencia como en investigación. El tercer apartado describe la actividad que se realiza actualmente, en la que la terminología ocupa un lugar central. En el cuarto apartado se hace una breve descripción del trabajo interdisciplinario que algunos miembros de la comunidad universitaria hemos llevado a cabo, tanto en docencia como en investigación, enmarcados en el área de la ingeniería lingüística. En el quinto y último apartado se establecen algunas conclusiones y se considera la necesidad actual de buscar mejores resultados a través del trabajo interdisciplinario. Al final del artículo, se listan los títulos que en el área de terminología han sido producidos en la Universidad de Los Andes.

La Universidad de los Andes

La hoy denominada Universidad de Los Andes (ULA) fue fundada el 29 de marzo de 1785 bajo la figura de una casa de estudios religiosos que creó el Obispo de entonces, Fray Juan Ramos de Lora. Esta casa fue luego bautizada con el nombre de Real Colegio Seminario de San Buenaventura de Mérida. Con el correr de los años, el seminario cambió de nombres y no fue sino hasta 1810 que fue convertido en Universidad.

La Universidad de Los Andes es una universidad nacional autónoma, financiada por el Gobierno en un 90%. Comprende un área de 407.276 m² distribuidos en los tres estados andinos, a saber: Mérida, Táchira y Trujillo. Además cuenta con extensiones en las ciudades de Guanare (Estado Portuguesa), Valle de La Pascua (Estado Guárico) y Barinas (Estado Barinas). Posee en la actualidad una matrícula estudiantil de 35.000 estudiantes de carreras de pregrado y de 9.000 estudiantes de programas de postgrado¹. La instrucción es impartida en idioma español, con un régimen de estudio por semestre, salvo en las carreras que se dictan en las facultades de Medicina y de Ciencias Jurídicas y Políticas.

La Universidad de Los Andes cuenta con diez facultades en la sede de Mérida y dos núcleos universitarios. Cada una de las facultades está integrada por escuelas, que a su vez se subdividen en departamentos; dependiendo de

las áreas de conocimiento a ellas adscritas, las cuales van desde las ciencias básicas, la ingeniería, la arquitectura, las ciencias del agro, las ciencias de la salud, las ciencias de la educación hasta las ciencias sociales, las humanidades y las artes. Cuenta con importantes institutos, centros y laboratorios de investigación, que sumados a la productividad académica de sus actores, sitúan a la Universidad de Los Andes como la segunda universidad del país.

La terminología: un poco de historia

Cuando en 1983 se llevó a cabo el Primer Seminario Nacional de Terminología en la Universidad Simón Bolívar de Caracas, una de las personas signatarias del germen, que después se convertiría en la Red Iberoamericana de Terminología (RITerm), fue Mary Castañeda (Fedor y Boroni, 1983). Esta docente de la Universidad de Los Andes, hoy fallecida, enseñaba principios teóricos y prácticos de la terminología acordes con la Teoría General de la Terminología, dentro del curso de traducción jurídica y legal correspondiente al programa de Licenciatura en Letras mención Lengua y Literatura Inglesas. Tanto en los tres cursos de traducción² correspondientes a la carrera de la licenciatura, como en el Servicio de Traducción se trabajaba en terminología, no tanto por estar al día con las corrientes teóricas del momento, sino por la necesidad externa que planteaba e impulsaba el trabajo terminológico.

El Servicio de Traducción de la Universidad de Los Andes fue creado oficialmente en 1974 con la finalidad de prestar servicio de traducción a la comunidad universitaria. Ya desde 1966, profesores adscritos al entonces Departamento de Idiomas Modernos (hoy convertido en Escuela de Idiomas Modernos), realizaban traducciones de artículos escritos en inglés, francés y alemán en sus ratos libres. Poco a poco la necesidad de sistematización, el aumento de personal y la dedicación al servicio fueron aumentando. Se contrataron traductores para los idiomas inglés y francés. Se iniciaron los cursos de traducción dentro del programa de la licenciatura del Departamento de Idiomas Modernos y, con ello, se creó la necesidad de elaborar vocabularios técnicos, en un intento por sistematizar y agilizar el trabajo del traductor.

Así pues, después de varios años recopilando información a partir de artículos científicos que llegaban al servicio, se publicaron dos trabajos, *Léxico en la construcción de puentes de concreto*, de Alba Linares de Rondón (Linares, 1981) y *Terminología estadística inglés-español*, de Yolanda Rincón de Rada (Rincón, 1981). El área de mayor

demanda, sin embargo, era el académico-legal³, por cuanto en el servicio se realizaban traducciones de los documentos académicos requeridos por los profesores de la universidad que salían o llegaban del extranjero por beca o año sabático.

Obviamente, la terminología legal y educativa siempre ha sido motivo de preocupación de parte de los traductores por cuanto muchos documentos se repetían en forma y contenido y había necesidad de agilizar de alguna forma el trabajo repetitivo de este tipo de documentos. Por otra parte, como es bien sabido, los sistemas educativos tanto de Venezuela como de los Estados Unidos e Inglaterra (países de mayor preferencia para realizar estudios de postgrado o pasantías de investigación) son bastante disímiles. La recopilación, uso y control de una terminología coherente para la traducción de los documentos se convirtió en la piedra angular del trabajo de varios profesores del área de la traducción. De esta manera, en 1988 se publicó el Glosario de términos educativos inglés-español de Mary Castañeda, el cual hoy en día se sigue consultando con bastante asiduidad y se utiliza como referencia obligada en el curso de traducción de documentos académicos y legales.

Por años, Mary Castañeda tuvo la idea de crear el Banco de Datos de la Universidad de Los Andes (BDTULA), cuyo ambicioso fin era centrar la terminología de las diversas áreas de conocimiento de la universidad mediante la alimentación de un sistema informatizado que fuese manejado por personal dedicado a tal fin. En vista de la inviabilidad de manejar semejante cúmulo de información especializada, se optó por el diseño de un banco de datos del Servicio de Traducción, cuyo fin era registrar y controlar la terminología utilizada en los documentos educativos y legales que se procesaban en el Servicio de Traducción.

Después de varios años de reuniones, vaciado manual de la información terminológica y acuerdos en el diseño de la ficha de trabajo, se contrataron los servicios de un informático para diseñar el programa bajo la plataforma DOS, mediante el programa que para entonces era de libre distribución: el ISIS. De esta manera, desde 1993 hasta 1997 se ingresaron en el BDTULA más de 400 fichas bilingües de términos en el área académico-legal. Lamentablemente, hoy en día se desconoce la ubicación del BDTULA, puesto que era de consulta restringida y de un usuario a la vez. No se llegó a publicar en la red Internet, a pesar de que la Universidad de Los Andes ha sido pionera en Venezuela, por ser la primera institución de educación superior que ya en 1991 contaba con una red de datos con salida al exterior. Del contenido de este programa queda afortunadamente el trabajo Desarrollo de una base de datos terminológicos para el servicio de traducción de la Universidad de Los Andes, publicado en

formato papel por Heberto Fernández (Fernández, 1995). En la Facultad de Ciencias Jurídicas y Políticas, se está elaborando un trabajo terminológico en el área del derecho electoral venezolano, coordinado por la profesora Ramos García de Méndez y asesorado por especialistas en el área del derecho electoral.

La terminología de otras áreas también ha sido motivo para la realización de interesantes trabajos. A lo largo de los años, los especialistas han dedicado enormes esfuerzos aislados para producir glosarios, vocabularios y diccionarios terminológicos en sus respectivas áreas, impulsados especialmente por la ausencia de material terminológico actualizado para responder a sus necesidades y a las de sus estudiantes. Así pues, la Universidad de Los Andes, cuenta con obras terminológicas de diverso formato producidas en áreas como zoología, matemáticas, agricultura y ciencias forestales, entre otras⁴.

La premisa del trabajo en tándem entre el especialista de la materia y el traductor se inició con el esfuerzo mancomunado del proyecto BDTULA, pero se hizo patente en la elaboración del Glosario terminológico de redes teleinformáticas inglés-español, desarrollado por Melva Márquez Rojas entre 1995 y 1996, bajo la tutoría de José Gregorio Silva, especialista en redes de datos y uno de los fundadores de la Red de Datos de la Universidad de Los Andes (RedULA), hoy convertida en modelo de red después de más de diez años de funcionamiento. El glosario contiene 196 términos clasificados temáticamente y extraídos de un corpus textual conformado por más de 15 textos originales, tanto en inglés como en español. Las dos lenguas del glosario (inglés y español) entran en contacto tan solo a través del campo de la equivalencia, puesto que es un glosario básicamente monolingüe. Por cada término se presenta, además, información de género gramatical y de los procesos de formación léxica involucrados, así como campos de definición, ejemplos de uso y las fuentes bibliográficas empleadas. El formato de presentación de la ficha terminológica siguió las recomendaciones establecidas en el texto de Cabré (1993). El glosario fue diseñado con el programa FoxPro® para Windows y ha estado en la red Internet desde diciembre de 1996 hasta el presente, en la dirección <http://www2.adm.ula.ve/sister>. Sin embargo, la falta de recursos financieros ha impedido hasta el momento mantener actualizado el sistema, llamado Sister (por "sistema terminológico"), a pesar de que se mantiene vigente en su servicio con más de 50 consultas diarias recibidas en las distintas áreas de la red.

La terminología hoy

El trabajo en terminología no se detiene, como tampoco desaparecen los obstáculos para llevarlo a cabo. No obs-

tante, los esfuerzos de algunos profesores del área por contactar especialistas en terminología, realizar cursos de postgrado en el área y adquirir bibliografía adecuada y reciente, han dado buenos resultados ya que se ha despertado en la comunidad estudiantil el interés por trabajar no sólo en terminología, sino en documentación y traducción asistida por computador.

Hoy día, la Escuela de Idiomas Modernos cuenta con la opción "traducción" en la licenciatura en Idiomas Modernos, además del Servicio de Traducción. Las lenguas de trabajo son el inglés y el francés y los cursos son dictados en traducción directa, no inversa. En el programa de la carrera se ofrecen cinco cursos de la opción, a saber, Traducción I (historia y teoría de la traducción, principios básicos del proceso traductológico, técnicas de traducción), Traducción II (traducción de textos científicos y técnicos, lenguajes especializados), Traducción III (traducción de textos legales y académicos), Traducción IV (traducción de textos literarios—desde 1997 hasta junio 2001- y la traducción en las nuevas tecnologías—septiembre de 2001 hasta septiembre de 2002) y el curso optativo Documentación y Terminología.

Desde el punto de vista de la investigación y producción de recursos terminológicos, la terminología se encuentra estrechamente vinculada con el trabajo interdisciplinario que en el siguiente apartado se describirá.

La ingeniería lingüística: necesidad interdisciplinaria

Por todos es bien sabido que la ingeniería lingüística es una disciplina relativamente nueva, pero pujante, que se nutre de dos áreas de conocimiento: la computación y la lingüística. Su objetivo principal es la aplicación del conocimiento de la lengua al desarrollo de sistemas informáticos capaces de reconocer, comprender, interpretar y generar el lenguaje humano en todas sus formas. En los últimos años, se tiene conocimiento del desarrollo de sistemas automáticos y semiautomáticos que se orientan principalmente hacia la descripción de la lengua y hacia la extracción de información con fines muy diversos, que van desde la creación de diccionarios y el diseño de cursos de lenguas hasta la traducción automática y sistemas de autoaprendizaje en áreas del conocimiento especializado.

En el seno del Centro de Investigación y Proyectos en Simulación y Modelos (Cesimo), <http://cesimo.ing.ula.ve>, se creó recientemente el Grupo de Investigación en Ingeniería Lingüística (GIL), cuyas características más destacadas son la interdisciplinaria (el grupo está conformado por lingüistas, informáticos y matemáticos), el

origen pedagógico (el grupo está asociado al programa de maestría en Modelado y Simulación de Sistemas) y sus objetivos tecnológicos (procesamiento del lenguaje natural y desarrollo de productos informáticos con fines lingüísticos).

Además de la creación del grupo GIL, en la Universidad de Los Andes se ha incrementado últimamente la necesidad de trabajar mancomunadamente en torno a la solución de problemas propios de la diversidad conceptual presente en la institución. De esta manera, especialistas de las áreas de bioinformática, genómica e inmunología han dado a conocer sus necesidades de informar e informarse. Por ello, ha surgido el proyecto Bioinformantes (Dávila et al, 2001), es decir, bioinformación gestionada por agentes en Internet, y algunos temas en tesis de pregrado y postgrado cuyos objetivos comunes son la automatización de la información a través de las herramientas propias de la computación y la informática.

Actualmente se encuentra en etapa de diseño el proyecto Portal inteligente para documentación en genómica, cuyos objetivos son:

- crear un portal temático WEB, concentrado en el área de la genómica, que funcione como un centro virtual de aprendizaje y consulta para usuarios expertos y no expertos en biotecnologías;
- crear colecciones electrónicas de textos especializados producidos en el país y asociarles las estructuras (ontologías) pertinentes para la documentación de fácil aprendizaje y para la recuperación efectiva de información vía web y correo electrónico;
- construir y adaptar herramientas de software para indexación y clasificación automática de contenidos en genómica y áreas relacionadas, así como para la generación automática de resúmenes a partir de los documentos que existen en varios idiomas; y
- diseñar interfaces para la extracción de información que faciliten la interacción humano computador al combinar elementos gráficos con formas de lenguaje natural (del español), reconocibles por el computador.

Asimismo, se están creando programas dirigidos a la formación de humanistas en herramientas computacionales, como por ejemplo, el seminario Traducción y nuevas tecnologías, correspondiente a la asignatura Traducción IV que se dicta actualmente a estudiantes cursantes del último semestre de la opción Traducción en la Licenciatura en Idiomas Modernos de la Universidad de Los Andes. En este seminario se busca resolver problemas interlingüísticos auténticos expuestos por los especialistas, a través de la traducción de documentos, de la elaboración de vocabularios y glosarios terminológicos y de la edición textual de los documentos que los mis-

mos especialistas entregan. A modo de ejemplo, en el segundo semestre de 2001, los resultados de este seminario fueron un glosario bilingüe (inglés-español) sobre la influenza, un inventario de recursos terminológicos monolingües y bilingües en el área del genoma humano y un análisis de los formantes grecolatinos en la terminología de la malaria. Todo este material está siendo sometido al análisis de los especialistas para darle su debida difusión. En el presente semestre los estudiantes y la docente están trabajando en un glosario monolingüe sobre la malaria dirigido a niños en edad escolar, así como en el inventario de páginas dedicadas a la difusión de la malaria en países de habla hispana ubicados en las zonas de mayor riesgo de esta enfermedad endémica.

Otro ejemplo de este tipo de trabajo que acerca humanistas al uso de herramientas computacionales es el curso Procesamiento lingüístico de información especializada que actualmente se dicta como tópico especial dentro del programa de Maestría en Modelado y Simulación de Sistemas. Por ello, en este curso se pretende iniciar al estudiante en el estudio de la documentación, la terminología como interdisciplina y la lingüística de corpus, teniendo como base únicamente el lenguaje especializado de textos escritos. Para llevar a cabo la tarea nada fácil de intentar procesar el lenguaje humano en cualquier forma, es necesario proveer al estudiante formado en ingeniería de ciertas herramientas teóricas y prácticas que le faciliten la comprensión de su actividad y lo vinculen con la ciencia de la lingüística.

En el año 2001 se realizó con éxito el curso Lingüística Computacional dentro de este programa de postgrado. Estos dos cursos han marcado pauta en este programa interdisciplinario porque se ha reconocido el uso consciente del lenguaje como eje transversal en tópicos cuyos contenidos no se destacaban precisamente por la necesidad de vehicular el conocimiento a través del lenguaje.

Un ejemplo digno de mencionar dentro de este apartado es el curso Agentes, minería de datos y gestión del conocimiento: herramientas de bioinformática, que dictan en el programa de la Maestría en Computación Jacinto Dávila Quintero y Raúl Isea, cuyo objetivo principal es conocer, analizar y desarrollar herramientas computacionales empleadas para el modelado y gestión del conocimiento biológico, en particular el genético. En este curso se integran biólogos y computistas para el manejo del conocimiento, área muy cercana con el trabajo que se realiza tanto en la terminología como en la documentación.

Además de las actividades realizadas en torno a la investigación y docencia de áreas vinculadas con la terminología, es preciso mencionar que el Consejo Universitario de

la Universidad de Los Andes ha aprobado recientemente una propuesta de convenio presentada, a través del Cessimo, por Melva Márquez Rojas y Jacinto Dávila ante el Institut Universitari de Lingüística Aplicada de la Universidad Pompeu Fabra de Barcelona. Este convenio de cooperación tiene, entre sus objetivos, el intercambio de profesores y estudiantes de postgrado, el intercambio de información relacionada con temas de investigación, libros, publicaciones y otros temas de interés para ambas instituciones, y la realización de proyectos de investigación especialmente enmarcados en los ámbitos de la ingeniería lingüística, la lingüística computacional, la terminología, la neología y la bioinformática.

Conclusiones

En el presente trabajo se expusieron brevemente algunas de las actividades más relevantes estrechamente vinculadas a la terminología y los lenguajes especializados, que se realizan en la Universidad de Los Andes, universidad autónoma del Estado, ubicada en el centro de la región andina de Venezuela: Mérida.

En primer lugar, dimos a conocer el contexto académico institucional que rodea el trabajo docente e investigativo relacionado con la terminología y la ingeniería lingüística. Seguidamente, realizamos un breve recorrido por la historia del tratamiento que se le ha dado al estudio de la disciplina de la terminología dentro de los cursos de la opción Traducción del programa de licenciatura en idiomas modernos que se ofrece en la Facultad de Humanidades y Educación de la universidad. Posteriormente, describimos brevemente la situación actual de la terminología, con mayor participación dentro de las actividades de docencia e investigación. Por último, informamos sobre la necesidad sostenida, que hoy día se va satisfaciendo lentamente con la participación en proyectos conjuntos de especialistas en lengua, informática, lógica y biología. Esta actividad constituye para nosotros un estímulo que nos motiva en el trabajo interdisciplinario por el libre flujo del conocimiento que presenciamos y las oportunidades que se le ofrecen a los profesionales de intercambiar sus inquietudes para lograr objetivos comunes de cara al beneficio de la "sociedad de la información", término puesto en el tapete en los últimos años.

El conocimiento es producto de un proceso dinámico y, como tal, se fundamenta en gran medida de la transmisión de la información, principalmente a través de signos lingüísticos. En la Universidad de Los Andes, el dinamismo de la ciencia y la tecnología se traduce en la necesidad, manifiesta por sus principales mentores, de utilizar herramientas que puedan agilizar el libre flujo del conocimiento. No escapa a esta realidad el trabajo intensivo que traductores, hombres de ciencia y docentes han realiza-

do con miras a solventar el siempre existente desfase entre el concepto y su materialización a través del lenguaje.

Venezuela posee un territorio de gran fertilidad dentro de las lenguas de especialidad. Por ello, el trabajo de describir, recopilar, analizar, programar, modelar y difundir es muy arduo. El GIL es el primer grupo de investigación académico venezolano que intenta unificar esfuerzos, que hasta ahora se han realizado de forma aislada, en disciplinas como la terminología, la traducción, la informática, la biología, la lógica y la documentación. La creación de este grupo constituye el punto de partida de un tímido desarrollo que busca incentivar la docencia y la investigación en áreas troncales donde tanto la terminología, la documentación y la informática están ubicadas en el centro de sus actividades.

Melva J. Márquez R.



Jacinto A. Dávila Q.



Centro de Investigación y Proyectos
en Simulación y Modelos
Universidad de Los Andes, Mérida, Venezuela

Notas

1. Datos suministrados por la Dirección de Servicios de Información Administrativa de la Universidad de Los Andes (www.ula.ve/dsia)

2. Traducción I (teoría de la traducción, etapas del proceso traductivo, técnicas de traducción), Traducción II (traducción de textos científicos y técnicos) y Traducción III (traducción de textos legales y académicos).

3 Esta área temática sigue siendo la de mayor demanda hoy en día.

4 Véase el apartado de la lista de "Productos terminológicos y lexicológicos de la Universidad de Los Andes" al final del artículo.

5 Este repertorio de recursos lexicológicos y terminológicos fue obtenido a través de la página web de los Servicios Bibliotecarios de la Universidad de Los Andes (<http://www.serbi.ula.ve>). Por lo tanto, el repertorio puede no ser exhaustivo ya que no todos los títulos de las colecciones académicas están indexados en la base de datos de acceso virtual.

Referencias

CABRÉ, MA. T. (1993). *La terminología. Teoría, metodología, aplicaciones*. Barcelona, España: Antártida.

CASTAÑEDA, M. (1988) *Glosario de términos educativos inglés-español*. Mérida: Universidad de Los Andes (trabajo de ascenso no publicado)

DÁVILA, J., A. BARRIOS Y J. LÓPEZ (2001) *Bioinformants. ELPCAR II* (Escuela Latinoamericana de Paralelismo y Computación de Alto Rendimiento). Mérida: Universidad de Los Andes.

FEDOR, A. Y BORONI, M. (Eds) (1984) *Actas del primer seminario nacional de terminología*. Caracas: Universidad Simón Bolívar

FERNÁNDEZ U., HEBERTO H. (1995) *Desarrollo de una base de datos terminológicos para el servicio de traducción de la Universidad de Los Andes*. Mérida: Universidad de Los Andes (trabajo de ascenso no publicado)

LINARES DE R., A. (1981) *Léxico en la construcción de puentes de concreto*. Mérida: Universidad de Los Andes (trabajo de ascenso no publicado)

MÁRQUEZ, M. (1996) *Glosario terminológico de redes teleinformáticas inglés-español*. Mérida: Universidad de Los Andes (trabajo de ascenso disponible en <http://www2.adm.ula.ve/sister>)

Productos terminológicos y lexicológicos de la Universidad de los Andes⁶

1. BUSCHI, ALFREDO (1974) *Vías de comunicación*. Mérida: Universidad de Los Andes.

2. CARABALLO CAMPOS, JOSEFINA (1995) *Glosario de zoología*. Mérida: Universidad de Los Andes, Consejo de Publicaciones (ISBN : 980-221-777-8)

3. CASTAÑEDA, B., MARY (1988) *Glosario de términos educativos inglés-español*. Mérida: Universidad de Los Andes (trabajo de ascenso no publicado)

4. DEPARTAMENTO DE IDIOMAS-SERVICIO DE TRADUCCIÓN (1999) *Boletín de títulos traducidos* (inglés-español / español-inglés). Mérida: Universidad de Los Andes, Consejo de Publicaciones.

5. FERNÁNDEZ U., HEBERTO H. (1995) *Desarrollo de una base de datos terminológicos para el servicio de traducción de la Universidad de Los Andes*. Mérida: Universidad de Los Andes (trabajo de ascenso no publicado)

6. FERNÁNDEZ U. HEBERTO H. (1999) *Un aporte a la terminología de la didáctica de la traducción*. Mérida: Universidad de los Andes (trabajo de ascenso no publicado)

7. FLORES ORTEGA, BERNARDO ENRIQUE (1985) *El Léxico de la alimentación en el habla culta de San Cristóbal*. San Cristóbal: Universidad de Los Andes, Núcleo Universitario del Táchira.

8. FLORES ORTEGA, BERNARDO ENRIQUE (s.f.) *Gastronomía y léxico del habla culta en la frontera colombo venezolana*. San Cristóbal: Universidad de Los Andes, Núcleo Universitario del Táchira.

9. GONZÁLEZ RAPOSO, MARÍA DEL SALVADOR (s.f.) *Evaluación del vocabulario matemático en el ciclo básico común*. Mérida: Universidad de Los Andes

10. GRESBACH, JEAN CLAUDE (1972) *Cien términos usuales en geomorfología*. Mérida: Universidad de Los Andes, Facultad de Ciencias Forestales.

11. GRUPO DE INVESTIGACIONES SOBRE AGRICULTURA, GERENCIA Y AMBIENTE (2000) *Glosario de términos*. Mérida: Universidad de Los Andes, Consejo de Publicaciones (980-11-0370-1)

12. LÁREZ VELÁSQUEZ, CRISTÓBAL JOSÉ (2001) *Resumen de terminología básica utilizada en polímeros*. Mérida: Escuela Internacional de Polímeros, Universidad de Los Andes

13. LINARES DE RONDÓN, ALBA (1981) *Léxico en la construcción de puentes de concreto*. Mérida: Universidad de Los Andes (trabajo de ascenso no publicado)

14. MÁRQUEZ, MELBA (1996) *Glosario terminológico de redes teleinformáticas inglés-español*. Mérida: Universidad de Los Andes (trabajo de ascenso disponible en <http://www2.adm.ula.ve/sister>)

15. MORADOR DE WETTSTEIN RAQUEL (1993) *Terminología operativa en ciencia política*. Mérida: Universidad de Los Andes, Facultad de Ciencias Jurídicas y Políticas (ISBN: 980-221-670-4)

16. PAREDES BRICEÑO, CÉSAR (1990) *Glosario etimológico para el estudio de ciencias de la salud*. Mérida: Universidad de Los Andes (ISBN: 980-221-337-3)

17. QUIJADA ROSAS, MARCELINO (1992) *Glosario de términos de genética, fitogenética y afines con especial énfasis en la mejora de árboles*. Mérida: Universidad de Los Andes (980-221-430-2)

18. QUINTERO MORENO, JUAN PEDRO (1985) *Diccionario enciclopédico*. Mérida: Universidad de Los Andes, Facultad de Ciencias Jurídicas y Políticas

19. RINCÓN DE RADA, YOLANDA (1981) *Terminología estadística inglés-español*. Mérida: Universidad de Los Andes (trabajo de ascenso no publicado)

20. RINCÓN DE VILLALOBOS, BELKYS M. (s.f.) *Glosario de la terminología de expresión gráfica plástica para preescolar y básica*. Mérida: Universidad de Los Andes.

21. RODRÍGUEZ CARRASQUERO, HENRY A. (1973) *Nombre científico y vulgar de 20 especies forestales de Venezuela*. Mérida: Universidad de Los Andes, Facultad de Ciencias Forestales

22. ZAPATA BECERRA, ARGENIS ARTURO (1985) *A Glossary of English Faux Amis for English Speakers*. Lawrence, Kansas: University of Kansas.

Introducción a la lingüística computacional con una perspectiva interdisciplinaria

Cet article présente un projet de cours d'introduction à la linguistique informatique élaboré par une équipe interdisciplinaire composé par des linguistes et des informaticiens. Il se base sur l'hypothèse que des études de ce type peuvent contribuer au renforcement de la formation, aussi bien des techniciens que des humanistes, pour le travail de recherche en linguistique générale et en ingénierie de contenus en particulier. Le cours est fondé sur des exercices de logique informatique mais sont cependant destinés à supporter la discussion sur des thèmes spécifiques et généraux de la linguistique. Nous disposons, comme complément opératif, d'outils adaptés au traitement automatique de l'espagnol, tels qu'un syllabeur et un résumeur de textes.

Este artigo apresenta um projeto de um curso de introdução à lingüística informática, elaborado por uma equipe interdisciplinar composta por lingüistas e informáticos. Ele se fundamenta na hipótese que os estudos deste tipo podem contribuir para o reforço da formação, tanto dos técnicos como dos humanistas, para o trabalho de pesquisa em lingüística geral, e para a criação de conteúdos em particular. Embora o curso se baseie em exercícios de lógica informática, eles são destinados para apoiarem a discussão sobre temas específicos, e gerais, da lingüística. Como complemento operacional, ele dispõe de ferramentas adaptadas ao tratamento automático do espanhol, tais como um silabador e um resumidor de textos.

Introducción

El presente artículo muestra el diseño de un curso introductorio a la lingüística computacional producido por un equipo interdisciplinario conformado por lingüistas¹ y computistas². La contribución es un plan de estudio que se apoya en herramientas computacionales para mostrar al estudiante conceptos lingüísticos de cara al procesamiento del lenguaje natural. Nuestra hipótesis es que planes de estudio similares pueden contribuir a fortalecer la formación, tanto de técnicos como de humanistas, para el trabajo de investigación en lingüística general y en ingeniería de contenidos en particular.

El artículo está organizado de la siguiente manera: en primer lugar, explicamos los objetivos y el plan general del curso que se ha diseñado. A continuación, describimos las herramientas computacionales y los formalismos asociados que usamos en el curso e ilustramos los resultados de la aplicación de esas herramientas para abordar problemas de la lingüística en el español, la morfología y la extracción de información. En este aspecto, destacamos que los programas fueron diseñados por el equipo interdisciplinario, varios de cuyos participantes no tenían experiencia previa en programación. En la siguiente sección se presentan algunos conceptos de gramática, por entender que esta representa el núcleo duro de la lingüística [6] y que, a nuestro juicio, constituye la base para introducir a los participantes en la descripción de la lengua a través del análisis computacional. Posteriormente, se ilustra la flexibilidad del forma-

lismo computacional, mostrando una adaptación del célebre ejemplo de Eliza³, de un programa que manipula palabras y oraciones en inglés, a un programa que reconoce frases en español. La sección siguiente resume la discusión que el curso plantea en torno a los llamados metalenguajes. La última sección concluye con algunas referencias al plan de investigación que alojará el curso que hemos diseñado.

Presentación de la asignatura

El objetivo general del curso diseñado es conocer tecnologías de procesamiento del lenguaje natural escrito y aplicarlas al español y a las variantes locales de esta lengua. Los integrantes del grupo pretenden que estas tecnologías puedan aplicarse a objetivos tan diversos como estudios de la inteligencia y de la competencia verbal, ejercicios artísticos con el lenguaje, la explotación de recursos informáticos en el ámbito de las ciencias sociales y las humanidades y el modelado de sistemas cognitivos. Así, la asignatura está dirigida a estudiosos de los lenguajes (naturales o artificiales) que hayan tenido alguna exposición mínima a herramientas de computación y que quieran mejorar sus habilidades específicas.

La asignatura tiene los siguientes componentes: fundamentos y tópicos generales en lingüística. Los fundamentos son un recorrido introductorio por el procesamiento del lenguaje natural con Prolog. La adopción de este lenguaje de programación lógica obedece a: 1) Es

un lenguaje declarativo, el programador no está forzado a estructurar sus códigos como un algoritmo con un solo punto de inicio, una secuencia definida de comandos de ejecución y un punto de terminación. Por el contrario, los "textos" en Prolog son colecciones de cláusulas, reglas que "declaran" conocimiento y que permiten definiciones alternativas del mismo concepto; 2) Es un lenguaje computacional amigable con los no computistas: Los códigos Prolog están conformados por unos pocos símbolos elementales y mayoritariamente por los símbolos que introduce el programador. Quien codifica no está obligado a respetar una estructura rígida en el código, ni a declarar variables; 3) Es un standard industrial internacional, con realizaciones de bajo costo en todas las plataformas modernas de computación y disponible en forma gratuita para ejercicios académicos; y 4) Prolog es producto de los esfuerzos de investigación [5] que datan desde comienzos de los años 70 en el siglo XX, tanto en lógica y como en lenguajes formales. Esta tradición incluye varias disciplinas para representación del conocimiento y procesamiento del lenguaje natural.

De hecho, Prolog fue extendido desde sus inicios con un lenguaje de más alto nivel para el procesamiento del lenguaje natural, en particular para la descripción de cierto tipo de gramáticas generativas: *Definite Clause Grammar* [7], [5] o DCG, que presentan formalismos computacionales a los cuales pueden ser fácilmente inducidos los no computistas, en particular los estudiosos de la lingüística.

Otra ventaja del código Prolog es que puede ser empleado para verificar si cierta oración es gramatical según las reglas que se han definido.

Procesamiento del español en Prolog

Los programas constituyen un medio apropiado para guiar la discusión acerca de conceptos lingüísticos, manteniendo al mismo tiempo la referencia a una práctica que se puede verificar constantemente con ayuda del computador. Los programas generados en este curso están diseñados para procesar el idioma español en su forma escrita. El primero es específico para verificar la buena construcción métrica de los *haikus* (forma poética de origen japonés). Para ello, se usa un programa silabador del idioma español de alcance general que fue diseñado y programado durante las clases del curso. El segundo programa es un proyecto de investigación, llevado a cabo por algunos de los integrantes del grupo, sobre la extracción de información de textos especializados en español.

El *haiku* / silabador

Este programa descompone las oraciones en su estructura silábica y no considera las licencias poéticas. Ello requiere un tratamiento especial de la oración en el cual deben compararse sistemáticamente palabras contiguas, así como también debe dar cuenta de las licencias relativas a los hiatos. Se convierte el verso en una lista de sus sílabas y se verifica, a través de reglas escritas en Prolog, que se cumpla la regla métrica 5-7-5 exigida para un *haiku*.

A partir de la definición de un *haiku*, se transforma sistemáticamente un código directamente ejecutable en el computador. Luego, no sólo se obtiene el texto (por teclado o por archivo) del *haiku*, sino su análisis estructural. El silabador fue diseñado por nuestro equipo interdisciplinario para descomponer frases y oraciones, en sus sílabas constituyentes en español, a través del empleo de una definición de sílaba que hemos propuesto localmente.

El silabador descompone la oración en una lista de palabras, para luego fragmentar cada una de ellas en sus sílabas constitutivas. La primera operación es un procedimiento predefinido en la plataforma de computación Prolog, para leer el texto desde el teclado del computador y convertirlo en una lista de palabras. Hemos tomado la rutina propuesta por Bratko en [3].

El programa categoriza los distintos tipos de sílabas según los siguientes criterios:

- 1) Cuando la sílaba es terminal (incluye las aisladas);
- 2) La naturaleza de su núcleo vocálico; vocal simple, diptongo o triptongo.
- 3) Si el núcleo vocálico está precedido por una o dos consonantes como márgenes prenucleicos.
- 4) Relación del núcleo vocálico de una sílaba con el de la siguiente. Esencial para discriminar la partición de las consonantes entre núcleos vocálicos contiguos según las reglas morfológicas (márgenes pre y posnucleicos).

Una definición como ésta puede ser usada para verificar, incluir y crear instancias del lenguaje. Además, puede ser un buen medio para apoyar las discusiones y ejercicios intelectuales en procura de teorías lingüísticas generales o específicas a ciertos contextos. En ese sentido, nos parece que la herramienta informática no es recomendable sólo por su valor como generadora de lenguaje, sino también como medio de referencia para explicarlo. En consecuencia, en el curso insistimos en formar a los lingüistas tanto en el uso de las herramientas, como en su desarrollo y programación.

En las extensiones por venir, hemos incluido la creación de una herramienta interactiva para enseñar a los

niños a deletrear. Esta herramienta, que incluye una interfaz amigable para los niños, estará construida sobre el silabador y tendrá que desplegar la conducta interactiva típica de un instructor. Programada con el conocimiento del silabador y de cómo se aprende a deletrear y, además, con el objetivo de enseñar a deletrear, esta herramienta se comportará como un agente instructor.

El nivel de interacción de un instructor es posible con la tecnología actual de agentes inteligentes [18], restringiendo el "dominio del discurso" a un tema particular. El silabador podrá entender un rango de preguntas en español, siempre que se refieran al tema de deletrear. Al circunscribir el programa a ese dominio, problemas generales como referencias anafóricas y elípticas se simplifican. Debemos enfatizar, además, que esta tecnología es abierta, en tanto que está basada en un programa libre y en un lenguaje de representación del conocimiento bien fundado y ampliamente conocido: el lenguaje de la lógica.

El resumidor de textos en español⁴

Para complementar estos ejercicios de procesamiento del español, se tiene un código de computador, escrito en el lenguaje Prolog, que instrumenta un resumidor de textos basado en estilos [14]. Nuestra intención es que el uso de este programa contribuya a iluminar la discusión sobre la factibilidad de generar resúmenes automáticos de textos en lenguaje natural, en particular, en español, durante el curso de lingüística computacional. Los estudiantes tendrán oportunidad de conocer los detalles de una estrategia que hace factible el complejo proceso de elaborar resúmenes.

Si bien el código disponible se concentra en el análisis de oraciones y párrafos, con los criterios de claridad y coherencia de Williams [15], se le puede considerar una versión preliminar de un resumidor. La clave es usar la noción de tópico que propone Williams como guía para seleccionar los fragmentos relevantes del texto que se pueden incluir en un resumen.

Este programa opera sobre una representación "anotada" de cada oración según los criterios de Williams [15], con lo cual el programa separa sujeto, verbo y complemento en una revisión superficial de la gramática de cada oración ("superficial" implica la separación de tales componentes). Con esa revisión se identifica el verbo principal para dar paso al proceso de ajuste al contexto. Esta gramática se apoya en un diccionario de verbos (clasificados por categorías), para cuya construcción se empleó la herramienta TACT5 (*Textual Analysis Computing Tools*). La representación que emplea el

resumidor para procesar el texto es flexible; por lo tanto no se pretende un análisis sintáctico exhaustivo. Hemos podido extenderla para incluir un mecanismo de ponderación de importancia de cada tópico dentro del párrafo que, básicamente, aumenta la "prioridad" de cada tópico de acuerdo con las relaciones superficiales que interconectan gramaticalmente los componentes textuales. De esta manera se identifican la repetición de tópicos, las referencias anafóricas y algunos tipos de conexiones.

Un problema importante que debemos considerar en detalle en este proyecto es el concepto de relevancia o de un sustituto apropiado, respecto a algún contexto o solicitud de información. Los humanos pueden diferenciar la relevancia o la no-relevancia del contenido de los documentos respecto a cierta pregunta. Una noción de la relevancia ha sido propuesta por Cooper [16] y la denomina relevancia lógica. La importancia está definida en términos de la consecuencia lógica. Un documento es relevante a una necesidad de información, si y solamente si, contiene por lo menos una sentencia que sea relevante a esa necesidad.

Sin embargo, lo que se quiere es tener una estrategia genérica para extraer los tópicos relevantes independientemente de una necesidad específica de información. En otras palabras, los tópicos pueden ser relevantes o no dependiendo de la utilidad que se les dará, la cual es diferente para efectos de catalogación, terminología, procesamiento lingüístico o búsqueda de información. Una vez definida la utilidad del tópico será más sencillo discriminar la relevancia.

En vista de la importancia de la evaluación, será importante entonces considerar metodologías de evaluación de sistemas NLP, aunque estos pueden resultar muy costosos y complejos. Margaret King destaca la importancia de los resultados de las evaluaciones a sistemas de NLP como datos invaluableles, pero advierte que las evaluaciones varían enormemente en función del propósito, del alcance y de la naturaleza de los objetos que están siendo evaluados [17]. En nuestro caso, se seleccionó un dominio de conocimiento particular y se escogió un conjunto de artículos académicos de una revista periódica de circulación local. Esto nos ha permitido la interacción con algunos expertos en el dominio del contenido del texto para comparar la relevancia obtenida por el sistema.

Nuestro reto ahora es afinar la representación con criterios más sofisticados, sin desmejorar los logros que ya se obtienen. Creemos que esto es posible, pues la representación se presta a la elaboración y a la integración con otros mecanismos.

Introducción a las gramáticas contemporáneas

En este módulo de la asignatura se ofrece al participante del curso la exposición de conceptos fundamentales como la gramaticalidad [9] y la subcategorización verbal, así como una panorámica de las principales corrientes gramaticales que hoy día existen para describir y modelar el procesamiento del lenguaje natural, tales como la Gramática Léxico Funcional (LFG), la Gramática de Estructura de Frase Generalizada (GPSG) y, en especial, la Gramática de Estructura de Frase Regida por el Núcleo (HPSG). Este último modelo sintáctico y lexicalista toma de las otras dos gramáticas anteriores aspectos importantes para la descripción del lenguaje [8]; propone, además, la representación de las unidades lingüísticas como entradas léxicas en estructuras de rasgos de par atributo y valor que coindexan información léxica, sintáctica, semántica y pragmática. Con ello se pretende iniciar al participante en el reconocimiento de los formalismos lingüísticos utilizados en la lingüística moderna, de manera que pueda ser capaz de profundizarse en su estudio si decide tomar esta línea de investigación en su tesis de grado.

Elizaul: una introducción a la sociolingüística y a la pragmática

La relación del significado del signo con el contexto ha sido y sigue siendo uno de los problemas más complejos que ha enfrentado la lingüística y las disciplinas hermenéuticas en general. El problema consiste en describir el lenguaje como un fenómeno social que no tiene existencia independiente del hablante; se trata de desobjetivar la comunicación y reconocer el hecho intencional sobre el cual subyace. El proceso del descubrimiento del contexto ha sido lento en la larga historia de la cultura occidental: alrededor de unos 2500 años. Se ha pasado de la visión etimológica —de la búsqueda del significado original de las palabras— al concepto del uso lingüístico introducido por Ludwig Wittgenstein en sus Investigaciones Filosóficas [12]. En la antigüedad griega, los filósofos pensaban que las sonidos que constituyen las palabras están íntimamente ligados a la esencia de la cosa misma, es decir que el significado es absoluto e independiente de quien lo utilice.

Hoy en día nuestra visión de las cosas ha cambiado. Ya no esperamos entender el mundo a través de la morfología y fonética de las palabras. Hoy en día preferimos entender al ser humano a través del uso que hace éste del lenguaje y de los símbolos en general. Hemos descubierto tam-

bién, gracias a Austin [2] que con el lenguaje no sólo decimos cosas, sino que hacemos algo: hablando construimos nuestra identidad, cambiamos de status, nos comprometemos y hasta somos capaces de inventar mundos que aún no conocemos. El uso que hacemos del lenguaje nos permite indicar a qué grupo social pertenecemos, qué tipo de persona somos, en qué situación nos encontramos y de qué manera se debe interpretar lo que estamos diciendo. A cada situación le corresponde un tipo particular de vocabulario y un cierto tono de voz. No se trata solamente del uso especializado del lenguaje, sino de los marcos que constituyen nuestra experiencia.

Un ejemplo muy ilustrativo de lo que es el marco lingüístico se puede observar en la ya celebre Eliza [13], una pieza de software basada en plantillas que pretende simular a un psicoanalista en ejercicio. Las plantillas son sensibles a ciertas palabras claves que invocan una frase previamente diseñada: si el usuario menciona la palabra mother, Eliza le responderá: tell me more about your family. Eliza está lejos de ser un software inteligente, si por inteligente se entiende la capacidad de comprender a nivel semántico los enunciados del diálogo, pero se adecua muy bien para el modelado y simulación de diálogos en un contexto determinado.

La coherencia de una conversación con un software de este tipo depende de que el usuario identifique correctamente el contexto y el papel que debe desempeñar en éste. Cuando se intenta bromear con Eliza, rompiendo con el marco contextual, se obtienen respuestas incoherentes.

Durante el diseño de nuestro curso se modificó una versión de Eliza hecha en Prolog y se adaptó a otro contexto. A la nueva versión, que se comporta como un psicoanalista un poco atípico, la hemos llamado Elizaul. El marco de la consulta psicológica se ha mantenido, pero se le ha introducido un léxico que no corresponde a la situación. Nuestro analista padece un trastorno del lenguaje llamado coprolalia que se caracteriza por un habla obscena compulsiva que, en condiciones normales, podría resultar chocante y que haría pensar al paciente necesitado de ayuda que su terapeuta está loco. En este ejercicio no se trata de simular una situación excéntrica, sino más bien de explorar los marcos contextuales a través de una transgresión. Esta técnica, que consiste en poner léxico propio de contextos sociales de poco prestigio en la boca de actores "respetables", ha sido utilizada en diversas tradiciones artísticas y tiene como objetivo demarcar los límites de los marcos lingüísticos y socioculturales por medio del contraste y la descontextualización.

Ejercicios similares pueden servir para explorar, simular y describir un gran número de actos performativos en los que por lo menos uno de los participantes cambia de status al finalizar el mismo. En el caso de Eliza, se espera que el diálogo lleve al alivio del paciente y que éste pase del estatus de enfermo al de sano. Elizaul, por otra parte, intenta transformar al interlocutor a través de la risa, llevándolo de un estado de marcadas restricciones sociales a uno de menor control social⁷. De la misma manera se podría usar esta técnica para el estudio de la confesión católica o de la consulta médica como actos performativos, tratando de capturar el léxico propio del contexto, los actores y sus patrones de interacción, para luego someterlos a un proceso de modelado que puede reproducir de manera verosímil un situación sociolingüística. Desde luego que esto está en línea con el ejercicio que planteamos en la sección anterior con el agente silabador. Allí contamos con una ventaja adicional: el contexto es más fácil de restringir.

Metalingüajes

Llamamos metalingüajes a las especificaciones para manipular el conocimiento racional y lingüístico, es decir, un conjunto de convenciones que norman la comunicación en un lenguaje. El metalingüaje es un lenguaje que trata sobre lenguaje(s). En la sección de los metalingüajes se le ofrece al participante una visión en conjunto de los estándares internacionales utilizados para representar textos en formato electrónico. Los lenguajes de marcaje que se introducen son el SGML (*Standard Generalized Markup Language*) y el XML (*Extended Markup Language*). En el caso del lenguaje SGML, se introduce la noción de tipo de documento y la Definición de Tipo de Documento (DTD), las cuales se utilizan para dar formato a documentos existentes escritos en SGML. Uno de los principales objetivos planteados bajo esta sección es la vinculación de los recursos informáticos con el procesamiento de corpora lingüísticos, de manera que de su funcionamiento no sólo puedan beneficiarse las investigaciones de lingüistas y computistas, sino todas aquellas pesquias que, de una u otra manera, estén relacionadas con la descripción del lenguaje natural.

Conclusiones

Hemos diseñado un curso para inducir a computistas y lingüistas a compartir sus conocimientos y aprovechar las tecnologías de la información para el estudio de los lenguajes naturales. Los fundamentos del curso son los ejercicios de lógica computacional. Sin em-

bargo, éstos están expresamente dirigidos a soportar las discusiones sobre tópicos específicos y generales de la lingüística.

Creemos que el estudio de los lenguajes debe ocurrir en un entorno en el que, a la par del rigor científico, se entienda que el lenguaje es y será "constantemente innovador, ilimitado, libre -al parecer- del control de estímulos externos o estados de ánimos internos, coherente y apropiado a las situaciones" [4]. Éste es el contexto final de este ejercicio interdisciplinario.

Jacinto A. Dávila

Luis Astorga

Melva Márquez

Hilda Y. Contreras

Jacobo Myerston

M. Marilú Parra

Centro de Investigación y Proyectos

en Simulación y Modelos,

Universidad de Los Andes,

Mérida (Venezuela)

Referencias

- [1] ALARCOS LLORACH, E.: *Gramática de la Lengua Española*, Espasa-Calpe. Madrid, 1994.
- [2] AUSTIN, J.L. *How to do things with words*, Oxford University Press, 1962.
- [3] BRATKO, I.: *Prolog, Programming for Artificial Intelligence* (2da ed), Addison-Wesley. Harlow, 1990.
- [4] CHOMSKY, N. *El lenguaje y los problemas del conocimiento*. Madrid: Visor.
- [5] COVINGTON, MICHAEL. *Natural Language Processing for Prolog Programmers*. Prentice Hall. Englewood Cliffs, NJ, 07632. 1994. ISBN 0-13-629213-5
- [6] LERAT, P. (1997)(Trad. Albert Ribas) *Las lenguas especializadas*. Barcelona: Ariel (título original: Les langues spécialisées, 1995)
- [7] PEREIRA, F. AND SHEIBER, S. *Prolog and Natural-Language Analysis*, Center for the Study of Language and Information, 1987.
- [8] POLLARD C., Y SAG, I. *SAG Head-Driven Phrase Structure Grammar*. Stanford, CA: CSLI Lecture Note Series, Center for the Study of Language and Information. (1993)
- [9] SELLS, P. (1985) *Lectures on Contemporary Syntactic Theories*. Stanford, CA: Center for the Study of Language and Information.

[10] TUSON, J. (Dir.) (2000) *Diccionari de Lingüística*. Barcelona: Bibliograf.

[11] VON KLOPP, ANA. *A Practical Introduction to Prolog and Computational Linguistics*. Lecture notes. Disponibles desde email: a.von.klopp@bangor.ac.uk.

[12] LUWIG WITTGENSTEIN, *Philosophische Untersuchungen, Werkausgabe, Frankfurt am Main, Surhkamp* 1984, pp. 255-581

[13] JOSEPH WEIZENBAUM, *ELIZA: A Computer Program For the Study of Natural Language Communication Between Man and Machine*. Communications of the ACM. Volume 9, Number 1 (January 1966): 36-35.

[14] DÁVILA, JACINTO. Y CONTRERAS, HILDA. *Una gramática de estilos para resumir textos en español*. En XVII Congreso de la Sociedad Española de Procesamiento de Lenguaje Natural (SEPLN'2002), 11-13 septiembre.

[15] WILLIAMS, J. 1990. *Style: Toward Clarity and Grace*. The University of Chicago Press. Chicago and London.

[16] COOPER, W.S. 1971. *A definition of relevance for information retrieval*. *Information Storage and Retrieval*, 7, 19-37.

[17] King, M. 1996. *Evaluating Natural Language Processing Systems*. Communications of the ACM. Enero, Vol. 39, No. 1.

[18] WOOLDRIDGE, MICHAEL J; JENNINGS, NICHOLAS R (Eds.) (1995b). *Intelligent Agents*. Springer-Verlag, Berlin. 407 pages.

Notas

1. Melva Márquez es profesora de traducción y terminología, con maestría en lingüística aplicada por la U.S.B. y candidata al doctorado en lingüística aplicada por la Universidad Pompeu Fabra de Barcelona, España; Jacobo Myerston es filólogo (Master of Arts) y M. Marilú Parra es licenciada en letras y literatura latinoamericana.

2. Luis Astorga es matemático, Hilda Contreras es computista y Jacinto Dávila es profesor de lógica computacional.

3. Eliza, el computador terapeuta, fue creado por Joseph Weizenbaum del MIT en 1966 [13].

4. Página con interfaz web para el resumidor <http://cesimo.ing.ula.ve/INVESTIGACION/PROYECTOS/GIL/resumidor.html>.

5. TACT es un sistema de recuperación y análisis textual sobre bases de datos textuales en idiomas europeos. Este sistema se comenzó a desarrollar como una iniciativa de cooperación hacia las Humanidades por parte de IBM y la Universidad de Toronto durante los años 1986-89 (<http://www.chass.utoronto.ca/cch/tact.html>).

6. Utilizamos el término marco en el sentido que le dio Erwin Goffman en su libro *Frame-Analysis, an essay on the Organization of Experience*, New York, 1974.

7. En este sentido ver Mary Douglas, *Social Control of Cognition; Factors in Joke Perception*, en *Man*, N. S. 3, 3, pp. 361-367.