

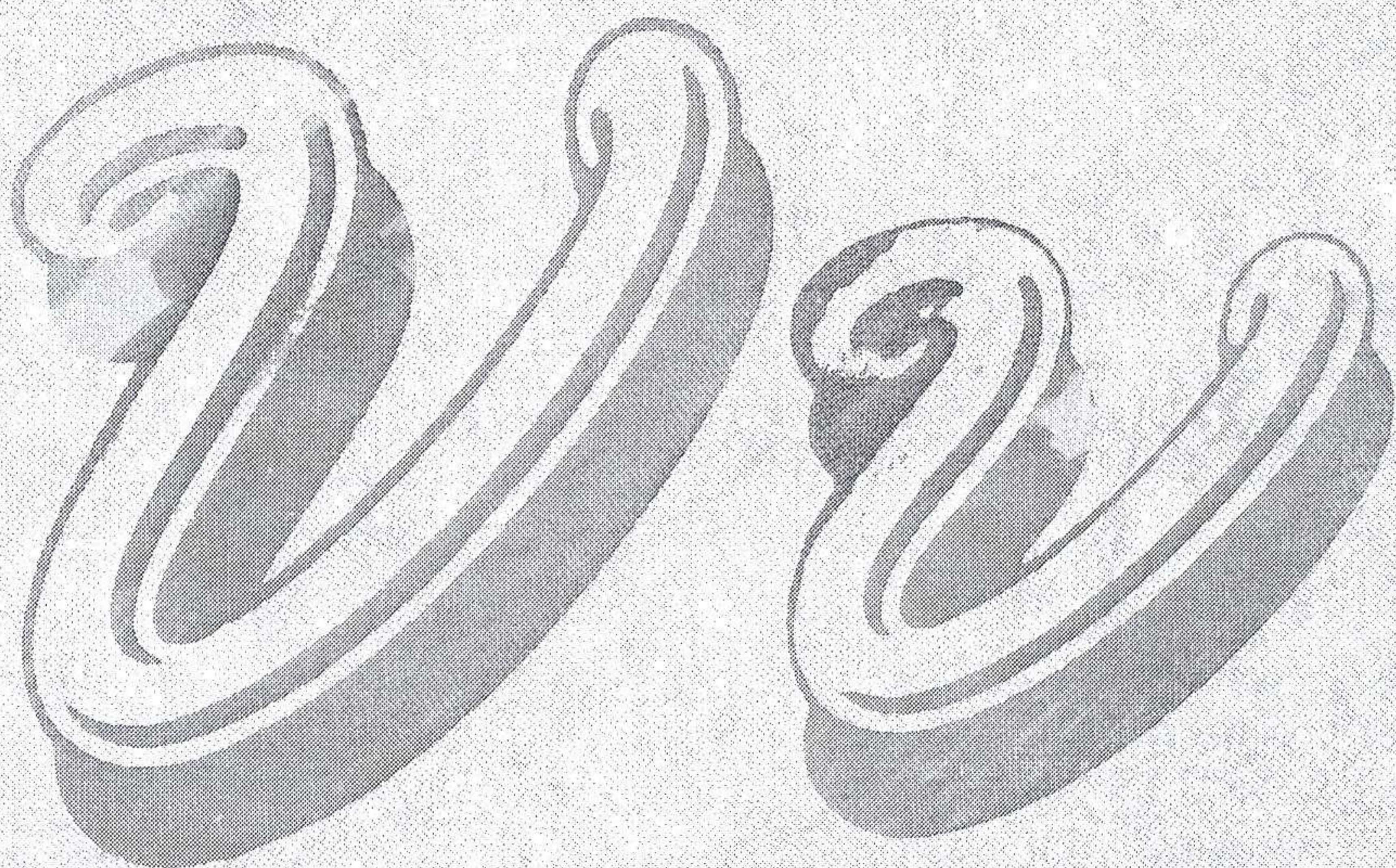
FUNDACIÓN GERMAN SÁNCHEZ RUIPÉREZ

Biblioteca del Libro

Las industrias de la lengua

Bajo la dirección de

José Vidal Beneyto



Hacia un doctorado europeo en lingüística computacional

ANTONIO ZAMPOLLI

Trad.: E. Lavín

Preámbulo

El campo de la Lingüística Computacional (LC) es tan amplio que aún no se ha intentado una definición completa generalizada de la disciplina.

La gran variedad de términos existentes para designarla en las diferentes lenguas confirma los intereses y aplicaciones varias que caracterizan a los investigadores que trabajan en este campo.

Creemos que sería útil resaltar algunos aspectos concretos:

— *La interdisciplinaridad de la LC*, debido a la contribución prestada de muchas maneras por diferentes disciplinas, localizadas, por lo general, en marcos didácticos y académicos que carecen parcialmente, por no decir completamente, de intercomunicación «administrativa».

— *La estrecha relación existente entre los aspectos prácticos*. La LC se ocupa de la aplicación de teorías lingüísticas a sistemas mecanizados de procesamiento del lenguaje natural dirigidos hacia finalidades específicas de «ingeniería» de las «industrias de la lengua», si bien su importancia intelectual es más grande de lo que se piensa.

«Sirve como foco de trabajo teórico para muchas disciplinas relacionadas entre sí, entre las que se incluye la lógica, filosofía, psicología, la ciencia teórica de la informática, y el habla y las comunicaciones. Este campo hace las veces de canal de influencia de doble dirección: el que parte de la teoría lingüística hacia las aplicaciones computacionales, y aquel otro que va desde las teo-

rías cognitivas y computacionales de los procesos del razonamiento hasta llegar a modelos de lenguaje y de su empleo.»

Multilingüismo

La dimensión multilingüe es un aspecto importante de la industria del lenguaje, sobre todo en Europa y en el marco actual de la llamada «sociedad de la información».

Tareas de investigación y de formación en esta área

La admisión de estudiantes en grupos experimentados de investigación es el instrumento más efectivo y, quizá, el más esencial para su formación y preparación en lingüística computacional durante el doctorado.

Sólo la colaboración entre universidades de varios países —en las que la investigación en el campo de la lingüística computacional está en marcha— puede ofrecer a los alumnos una serie de competencias, instalaciones, datos y actividades que cubran potencialmente la oferta completa de LC y la dimensión multilingüe necesaria.

Posibles materias

Los candidatos a los cursos de doctorado, normalmente:

- tienen estudios universitarios diferentes (Ital. «Facoltà»: lingüística, informática, disciplinas humanísticas, filosofía, psicología, etc.);
- necesitan especialización en una materia determinada de aplicación para la que se requiere una gran formación específica y fuertemente diferenciada: procesamiento del habla, traducción asistida, etc.

Hay que tener en cuenta los siguientes tipos de cursos:

a) *Requisitos y cursos de adaptación*

Parece esencial que los estudiantes empiecen los cursos con un bagaje mínimo de conocimientos. Una vez evaluado su nivel, hay que facilitar al estudiante que cubra sus huecos académicos, debido a sus estudios universitarios anteriores, en los siguientes sectores:

- lingüística teórica,
- aspectos de informática que guardan relación con el procesamiento del lenguaje natural (NLP),
- lógica.

b) *Cursos básicos*

Las asignaturas siguientes parecen ser las más firmes candidatas para los cursos fundamentales:

- tratamiento del lenguaje natural,
- morfología,
- sintaxis,
- semántica,
- lógica y lenguaje natural,
- analizadores,
- generación de frases,
- lexicología y lexicografía,
- tratamiento de textos,
- pragmática,
- ciencia cognitiva,
- AI y NLP; KR (L),
- análisis del discurso,
- hardware, software, lenguajes de programación para NLP,
- tratamiento de textos e investigaciones literarias, lingüística histórica, etc.

c) *Áreas de especialización*

Por ejemplo:

- Procesamiento de señales, síntesis y reconocimiento del lenguaje oral.

- Lingüística contrastiva, MHT, HAMT, MT.
- Estadística, lingüística cuantitativa, estilométrica.
- Ofimática.
- BD lingüísticas, IRT, documentación.
- Diálogo, interfaces en lenguaje natural.
- LC y ayudas a los disminuidos físicos.

Posibles contribuciones de Pisa

A. Las siguientes actividades didácticas se llevan actualmente a cabo en Pisa:

A.1. Nivel de licenciatura (Facultad de Letras).

Curso de iniciación en lingüística computacional. Asignaturas en curso: tratamiento de textos lingüísticos y literarios; lingüística estadística; lexicología computacional y lexicográfica; gramática formal; analizadores y generadores.

Curso de lingüística aplicada. Asignaturas en curso: lenguajes formales y autómatas, formalizaciones gramaticales y técnicas de análisis.

Cursos monográficos de lingüística computacional. Iniciación a la informática para lingüistas: lenguajes de programación; prácticas en los datos de LC a ILC.

A.2. Cursos de posgrado.

Plan de estudios de LC para el doctorado en lingüística. Los cursos se organizan en seminarios y, principalmente, como una participación teórica/práctica en la investigación de los grupos de lingüística computacional que trabajan tanto en el Departamento de Lingüística de la Universidad como en el Instituto de Lingüística Computacional del CNR (Centro Nazionale delle Ricerche) (véase más adelante).

Curso de lingüística computacional impartido en la escuela de especialización en «Ciencias computacionales». Asignaturas ofrecidas: tratamiento de lenguajes naturales y representación del conocimiento.

Debe tenerse en cuenta que la Facultad de Informática fue la primera que se creó en Italia, y que gran parte de la investigación del CNR en la rama de la ciencia computacional (que engloba a unos 300 investigadores, distribuidos en tres institutos: CNUCE, IEI, ILC) se concentra en Pisa.

B. Ya que pensamos que la especialización en el doctorado debe tener lugar principalmente a través de la participación en grupos y actividades avanzadas de investigación, enumeraremos los sectores en los que Pisa puede ofrecer una preparación adecuada debido a sus estudios de investigación nacionales (Progetti Strategici) e internacionales (ESPRIT, EUREKA, EUROTRA, «LARPA», etc.):

- Tratamiento de textos literarios y lingüísticos.
- Lexicología y lexicografía computacionales.
- Comprensión de lenguaje natural y diálogo.
- Bases de datos lingüísticos.
- Lingüística histórica, lenguas clásicas y LC.
- Teoría de gramáticas y de analizadores.
- Representación del conocimiento.

C. El Instituto de Lingüística Computacional y la Universidad de Pisa son de sobra conocidos por haber organizado varias Escuelas de Verano en lingüística computacional, cursos intensivos interdisciplinarios de tres a cuatro semanas de duración, que han contribuido en gran manera a la preparación de un buen número de investigadores que están trabajando en Europa y, parte, en los Estados Unidos.

El tema propuesto para la próxima Escuela de Verano, prevista para 1988 y financiada por la European Science Foundation, será lexicología y lexicografía computacionales.

D. *Instalaciones informáticas.* IBM 3081/K; varias impresoras láser y plotters; SUN 3/185 con ETHERNET; varias estaciones de trabajo (terminales, PCs, etc.); IBYCUS con CD-ROM.

E. *Datos y sistemas lingüísticos computacionales disponibles:*

- Procedimientos de procesadores de textos.
- Sintaxis computacional: formalizadores gramaticales y analizadores activos en lenguaje natural. Modelos de diálogo teóricos y computacionales.

- Lisp, PROLOG para tratamiento del lenguaje natural.
- Analizador morfológico computacional (italiano, español, latín).
- Lenguajes de representación del conocimiento.
- Gramáticas del italiano.
- Base de datos textual (50.000.000 palabras en 30 idiomas) y procedimientos de acceso.
- Diccionarios de máquina y bases de datos léxicos (mono y bilingües).