UNITED STATES PATENT AND TRADEMARK OFFICE

_____

BEFORE THE PATENT TRIAL AND APPEAL BOARD

_____

GOOGLE LLC,
Petitioner

v.

DIALECT, LLC,
Patent Owner

_____

Case IPR2024-00748
U.S. Patent No. 8,015,006 B2

DECLARATION OF INGRID HSIEH-YEE, PH.D.

Table of Contents

I. Introduction 3
A. Qualifications and Professional Experience 3
II. Background of Library Cataloging 4
A. Machine-Readable Cataloging ("MARC") 4
B. Library Online Catalogs 7
C. Scope of This Declaration 10
D. Evidence Considered in Forming My Opinion 13
III. Authenticity and Public Availability of *Zadrozny* (Ex. 1006) 16
A. Authentication 16
B. Penn Libraries Records 19
C. Harvard University Library Records 21
D. Internet Archive Records 24
E. Usage Record 25
F. Summary of My Opinion on *Zadrozny* 26
IV. Authenticity and Public Availability of *Gibbon* (Ex. 1008) 26
A. Authentication 26
B. Library of Congress Records 28
C. Copyright Registration Record 32
D. Usage Record 34
E. Summary of My Opinion on *Gibbon* 34
V. Authenticity and Public Availability of *Sutcliffe* (Ex. 1016) 35
A. Authentication 35
B. University of Wisconsin-Madison Library Records 36
C. University of California Berkely Libraries Records 38
D. Usage Record 41
E. Summary of My Opinion on *Sutcliffe* 41
VI. Conclusion 43

I. Introduction

I, Ingrid Hsieh-Yee, Ph.D., do hereby declare as follows:

1. I have been retained as an independent expert on behalf of Google LLC ("Petitioner") for proceedings before the United States Patent and Trademark Office ("PTO") regarding the authenticity and public availability of certain prior art references in relation to U.S. Patent No. 8,015,006.

2. I am being compensated for my work in this matter at my customary hourly rate. I am also being reimbursed for any reasonable expenses associated with my work and testimony in this investigation. My compensation is not contingent on the results of my study, the substance of my opinions, or the outcome of this matter.

A. Qualifications and Professional Experience

3. My complete qualifications and professional experience are described in my academic curriculum vitae (**Appendix A**). The following is a brief summary of my relevant qualifications and professional experience.

4. I was a professor in the Department of Library and Information Science at the Catholic University of America for 32 years and was conferred the title of Professor Emerita in December 2022. I have experience working in an academic library, a medical library, and a legislative library. I hold a Ph.D. and a Masters in Library and Information Studies from the University of Wisconsin-Madison.

5. I am an expert on library cataloging and classification and have published two editions on this subject, *Organizing Audiovisual and Electronic Resources for Access*: *A Cataloging Guide* (2000 and 2006). I taught a variety of courses, including Cataloging and Classification, Advanced Cataloging and Classification, Organization of Internet Resources, Organization of Information, Metadata, Digital Content Creation and Management, Internet Searches and Web Design, Information Literacy Instruction, Advanced Information Retrieval and Analysis Strategies, and The Information Professions in Society. I am familiar with metadata schema design and implementation. In my teaching I covered the design and implementation of metadata in databases, search engines, digital repositories, digital libraries, and digital archives. I also covered how information organization affects the discovery and access to digital resources on the Internet. My research interests cover cataloging and classification, information organization, metadata, information retrieval, information architecture, digital collections, scholarly communication, social media, user interaction with information systems, and others.

II. Background of Library Cataloging

A. Machine-Readable Cataloging ("MARC")

6. I am very familiar with a library cataloging encoding standard known as the “Machine-Readable Cataloging” standard, also known as “MARC,” which became the national standard for sharing bibliographic data in the United States by

1971 and the international standard by 1973. MARC is the primary communications protocol for the transfer and storage of bibliographic metadata in libraries. Experts in my field reasonably rely upon MARC records when forming their opinions.

7. A MARC record consists of several fields, each of which contains specific data about the work. Each field is identified by a standardized, unique, three-digit code corresponding to the type of data that follows. **Appendix B** is a true and correct copy of Parts VII to X of "Understanding MARC Bibliographic: Machine-Readable Cataloging" (http://www.loc.gov/marc/umb/um07to10.html) from the Library of Congress that explains commonly used MARC fields. For example, the personal author of a work is recorded in Field 100, the title is recorded in Field 245, publisher information is recorded in Field 260, the physical volume and characteristics of a publication are recorded in Field 300, and topical subjects are recorded in Fields 650.

8. I am very familiar with library cataloging and processing practices and the relationships between OCLC (Online Computer Library Center) and library catalogs, and how libraries create and share their records to help the public discover and access information.

9. The Online Computer Library Center (OCLC) is the largest bibliographic network of the world, and its WorldCat database contains more than 554 million records from thousands of member institutions (many of which are

libraries of some type) in more than 100 countries. OCLC was founded in 1967 to promote and support library cooperation. According to the "Third Article, Amended Articles of Incorporation of OCLC Online Computer Library Center, Inc.," OCLC was created "to establish, maintain and operate a computerized library network and to promote the evolution of library use, of libraries themselves, and of librarianship, and to provide processes and products for the benefit of library users and libraries, including such objectives as increasing availability of library resources to individual library patrons and reducing the rate of rise of library per-unit costs, all for the fundamental public purpose of furthering ease of access to and use of the ever-expanding body of worldwide scientific, literary and educational knowledge and information" (source: https://www.oclc.org/content/dam/oclc/membership/articles-of-incorporation.pdf).

10. OCLC members can contribute original cataloging records in MARC to the system or derive cataloging records from existing records, an activity referred to as "copy cataloging." When an OCLC participating institution acquires a work, it can create an original MARC record for the work in OCLC's Connexion system (a system for catalogers to create and share MARC records), and the system will automatically generate a code for the date of record creation in the yymmdd format, and the creating library's OCLC symbol is recorded in subfield "a" of the 040 field.

Once the MARC record is in Connexion, it becomes available to other OCLC members for adoption to their local online catalogs (*i.e.*, copy cataloging).

11. After a MARC record is created in Connexion, it also becomes searchable and viewable on WorldCat, which is a free web portal for users to explore more than 10,000 libraries worldwide. Records in WorldCat, however, are not presented in MARC fields. Instead, the data elements are labeled to help users interpret the records. Thus, the information stored in MARC records in Connexion is available to the interested public through the user-friendly WorldCat web portal. WorldCat records can also be searched on OCLC FirstSearch, a subscription-based search system that provides access to many proprietary databases, including the WorldCat database.

B. Library Online Catalogs

12. Library online catalogs gained acceptance in the early 1980s and many libraries migrated their systems to the World Wide Web in the mid-1990s. Library online catalogs are based on MARC records that represent their collections in order to help the public understand what materials are publicly accessible in those libraries. Most libraries with online catalogs have made their catalogs freely available on the Web. These online catalogs offer user-friendly search interfaces. Strong user interest in keyword searches and the popularity of Google have led to the "googlization" of library search systems. As a result, many library catalogs

nowadays provide a single search box for users to conduct keyword searches, with additional support for searches by author, title, subject terms, and other data elements such as ISBN (International Standard Book Number). Library catalogs these days also offer features for users to narrow their search results by language, year, format, and other elements. Many libraries display MARC records on their online catalogs with labels for the data elements to help the public interpret MARC records. Many libraries also offer the option to display MARC records in MARC fields.

13. Libraries create MARC records for works they acquire, including books, serials, motion pictures, and publications in other formats. Monograph cataloging is fairly common in libraries, and most libraries make a newly cataloged monograph available to the public soon after the cataloging work is completed, usually within a week. Libraries can create original cataloging records or use an existing record in OCLC to create a copy cataloging record. As soon as the cataloging record is completed, it is added to the library's online catalog for users. If the record is an original record, it is also entered into the OCLC system. If it is a copy cataloging record, the library's holding symbol is attached to the existing original record in the OCLC system to facilitate searching and interlibrary loan.

14. I am personally familiar with many online catalogs, databases, and search engines. In preparing for this declaration, I used the following authoritative information systems to search for records:

- Google Scholar (https://scholar.google.com/)
- Internet Archive (https://archive.org)
- Online catalog of Harvard University Library (https://hollis.harvard.edu/primo-explore/search?vid=HVD2&sortby=rank&lang=en_US)
- Online catalog of the Library of Congress (https://catalog.loc.gov/)
- Online catalog of the Penn Libraries of the University of Pennsylvania (https://www.library.upenn.edu/)
- Online catalog of the University of California Berkeley Library (https://search.library.berkeley.edu/discovery/search?vid=01UCS_BER:UCB)
- Online catalog of the University of Wisconsin-Madison Library System (https://search.library.wisc.edu/search/system)
- Public catalog of the US Copyright Office (https://cocatalog.loc.gov)
- WorldCat (https://search.worldcat.org/)

C. Scope of This Declaration

15. I have been asked to offer an opinion on the authenticity and public availability date of the following documents:

1) *COLING 94 : the 15th International Conference on Computational Linguistics : proceedings : August 5-9, 1994, Kyoto, Japan*, **Ex. 1006** ("*Zadrozny*");
2) *Handbook of Multimodal and Spoken Dialogue Systems : Resources, Terminology, and Product Evaluation*, (2000), edited by Dafydd Gibbon, Inge Mertins, Roger K. Moore, Kluwer Academic Publishers, **Ex. 1008** ("*Gibbon*");
3) Sutcliffe, R. F. E., Koch, H., & McElligott, A., (1996), *Industrial Parsing of Software Manuals*, Rodopi, **Ex. 1016** ("*Sutcliffe*").

16. I am rendering my expert opinion on the authenticity of the documents referenced herein and on whether these documents were published. I will provide my opinion on when and how these documents were publicly disseminated or otherwise made publicly available to the extent that persons interested and ordinarily skilled in the subject matter or art, exercising reasonable diligence, could have located and retrieved these documents.

17. I am informed by counsel that an item is considered authentic if there is sufficient evidence to support a finding that the item is what it is claimed to be. I

am also informed that authenticity can be established based on the contents of the documents themselves, such as the appearance, content, substance, internal patterns, protocols followed or other distinctive characteristics of the item.

18. I am informed by counsel that a given reference is "published" upon a satisfactory showing that such a document has been disseminated or otherwise made publicly available to the extent that persons interested and ordinarily skilled in the subject matter or art could locate it exercising reasonable diligence and obtain the document. I have also been informed by counsel that materials available in a library constitute "printed publications" if they are catalogued and indexed according to general library practices and protocols that make the references available and accessible to members of the interested public.

19. I have been informed by counsel that the '006 patent was filed in the United States and it claims "June 3, 2002" as its earliest priority date. I have assessed public accessibility of the documents referenced below as of this claimed priority date.

20. I have been informed by counsel that the subject matter of this proceeding generally relates to speech- and natural language-interfaces for interacting with and directing computer systems and software, including speech recognition and processing.

21. I have been informed by counsel that a "person of ordinary skill in the art at the time of the invention" (POSITA) is a hypothetical person who is presumed to be familiar with the relevant field and its literature at the time of the inventions. I have been informed by counsel that this hypothetical person is also a person of ordinary creativity, capable of understanding the scientific principles and literature applicable to the pertinent field.

22. I have been informed by counsel that a person of ordinary skill in the art would have possessed a bachelor's degree in electrical engineering or computer science, and two years of work experience in the field of speech- and natural language-interfaces for interacting with and directing computer systems and software, including speech recognition and processing. A person could also have qualified as a person of ordinary skill in the art with some combination of (1) more formal education (such as a master's of science degree) and less technical experience, or (2) less formal education and more technical or professional experience.

23. It is my opinion that such a person would have been actively engaged in learning about the field, possibly through formal instruction through the bibliographic resources. By at least 2002 such a person would have had access to a vast array of print and electronic resources, including at least the documents referenced below.

D. Evidence Considered in Forming My Opinion

24. In the preparation of this declaration, I have reviewed the documents referenced below and any other documents I reference herein, and each of these is a type of material that experts in my field would reasonably rely upon when forming their opinions:

1) The document referenced above in Section II.C;

2) Library copy of *Zadrozny* obtained from Harvard University Libraries, **Appendix 1006A**;

3) Library copy *Zadrozny* obtained from the Penn Libraries of the University of Pennsylvania, **Appendix 1006B**;

4) Bibliographic and MARC records for *Zadrozny* available at https://franklin.library.upenn.edu/catalog/FRANKLIN_9919010503503681 from the online catalog of the Penn Libraries of the University of Pennsylvania, **Appendix 1006C**;

5) Bibliographic and MARC records for *Zadrozny* available at https://id.lib.harvard.edu/alma/990059738670203941/catalog from the online catalog of the library of Harvard University, **Appendix 1006D;**

6) Webpage of the online catalog of Franklin, the online catalog of the

Penn Libraries, archived by Internet Archive on January 10, 1998, available at https://web.archive.org/web/19980110012724/http://www.franklin.library.upenn.edu/; and webpage of Hollis, the online catalog of Harvard University archived on December 10, 1997, available at https://web.archive.org/web/19971210172035/http://hplus.harvard.edu/alpha/HOLLIS.html, **Appendix 1006E**;

7) Citations to articles included in *Zadrozny*, obtained from Google Scholar, **Appendix 1006F**;

8) Bibliographic and MARC records for *Gibbon*, available at https://lccn.loc.gov/00044780 from the online catalog of the Library of Congress, **Appendix 1008A**;

9) Copyright registration record for *Gibbon*, available at https://cocatalog.loc.gov/cgi-bin/Pwebrecon.cgi?Search_Arg=Handbook+of+multimodal+and+spoken+dialogue+systems+&Search_Code=TALL&PID=G3a1EI3aX-CsJ9I10reaXgwutEmxnOL&SEQ=20240313190328&CNT=25&HIST=1 from the public catalog of the U.S. Copyright Office,

Appendix 1008B;

10) Copy of the online catalog of the Library of Congress archived by Internet Archive on February 29, 2000, available at https://web.archive.org/web/20000229084601/http://catalog.loc.gov/, **Appendix 1008C**;

11) Citations to *Gibbon*, obtained from Google Scholar, **Appendix 1008D**;

12) Bibliographic and MARC records for *Sutcliffe*, available at https://search.library.wisc.edu/catalog/999805139302121 from the online catalog of the University of Wisconsin-Madison General Library System, **Appendix 1016A**;

13) Bibliographic and MARC records for *Sutcliffe*, available at https://search.library.berkeley.edu/permalink/01UCS_BER/1thfj9n/alma991065546599706532 from the online catalog of the University of California, Berkeley Library, **Appendix 1016B**;

14) Webpage of the "Library Online Catalog Systems" available at the University of California Berkeley Library and the rest of the University of California system archived by Internet Archive on June 7, 1997, available at

https://web.archive.org/web/19970607084855/http://www.lib.berkeley.edu/Catalogs/, **Appendix 1016C**;

15) Citations to *Sutcliffe*, obtained from Google Scholar, **Appendix 1016D**.

III. Authenticity and Public Availability of *Zadrozny* (Ex. 1006)

A. Authentication

25. **Exhibit 1006** is a true and correct copy of portions of *COLING 94 : the 15th International Conference on Computational Linguistics : proceedings : August 5-9, 1994, Kyoto, Japan*, volume II ("*Zadrozny*"), that I received from counsel. I understand that counsel obtained this copy from the Library of Congress. This copy is presented as **Ex. 1006** in this declaration.

26. Page 1 is the spine and covers of a bound volume, and the spine identifies this work as "International Conference on Computational Linguistics, COLING 94" and it has a call number of "P 98 .I55 1994 Vol. 2 Copy 1." Page 2 is the title page that shows the title of this work is "COLING 94 The 15th International Conference on Computational Linguistics, Proceedings, Vol II" and the conference took place on "August 5 – 9, 1994, Kyoto, Japan." This page carries a date stamp of "Library of Congress MAY 07 1998 Purchase." Page 4 is another copy of the title page without the date stamp. Page 5 shows a hand-written call number of "P98 .I55 1994 Vol. 2 Copy 2" and a hand-written number of "2006438697" that appears to

be the Library of Congress control number. Page 6 is the Preface, pages 7 to 12 are the table of contents of volume 1, and pages 13 to 18 are the table of contents of volume II that shows the Zadrozny article begins on page 1289 and is the last article in volume II. Pages 19 to 23 (internal pages 1289 to 1293) show the full text of the Zadrozny article. Page 24 is the back cover of volume II. **Ex. 1006** shows the Zadrozny article has a total of five pages, including 10 references.

27. To authenticate **Ex. 1006**, I searched WorldCat for libraries holding *Zadrozny*, and learned that Harvard University Library was one of the holding libraries. After I verified the holdings information in their online catalog, counsel obtained a copy of *Zadrozny* from this library for me. This Harvard University library copy is presented as **Appendix 1006A** in this declaration.

28. **Appendix 1006A** is a true and correct copy of portions of *Zadrozny* obtained from the Harvard University Library. Page 1 shows the title page, and page 2 shows a hand-written call number of "WID-LC P98 .I55 1994x Vol. 2" and a library date stamp of "Harvard University Library FEB 26 1996." Pages 3 to 8 are the table of contents of volume II, and pages 9 to 13 are the full text of the Zadrozny article. Page 14 is the back cover. I have closely compared the Harvard copy (**Appendix 1006A**) with **Ex. 1006** and found the title page, table of contents for volume II, and the full text of the Zadrozny article in both copies to be the same.

29. My research on WorldCat on *Zadrozny* found the Penn Libraries of the University of Pennsylvania was another holding library of this work. After I verified the holdings information in their catalog, I asked counsel to obtain a copy of *Zadrozny* from this library. This Penn Libraries copy is presented as **Appendix 1006B** in this declaration.

30. **Appendix 1006B** is a true and correct copy of portions of *Zadrozny*. It includes the title pages, table of contents for volume I and volume II, and the Zadrozny article (internal pages 1289 to 1293). I have closely reviewed this Penn Libraries copy and compared it with the Library of Congress copy (**Ex. 1006**) and the Harvard University Library copy (**Appendix 1006A**), and found that all the pages in **Appendix 1006B** are the same as their corresponding pages in **Ex. 1006,** and all the pages in **Appendix 1006B**, except for the table of contents of volume I, are the same as their corresponding pages in **Appendix 1006A**, which does not include the table of contents of volume I.

31. My comparison of **Ex. 1006**, **Appendix 1006A**, and **Appendix 1006B** informs my opinion that **Ex. 1006** contains the same content presented in **Appendix 1006A**, and **Appendix 1006B**. **Ex. 1006** is in a condition that creates no suspicion about its authenticity. Specifically, the text on pages included in **Ex. 1006** appears to flow seamlessly, and there are no visible alterations to the pages. Furthermore, **Ex. 1006** is in the custody of a national library, and its content is the same as the

copies held by two major universities. Based on my analysis, it is my opinion that **Ex. 1006** is an authentic copy of *Zadrozny* and I see no reason to question its authenticity.

B. Penn Libraries Records

32. The Library of Congress MARC record for *Zadrozny* (**Ex. 1006**) shows they completed their cataloging record in 2007. To determine when this proceedings first became publicly accessible, I reviewed library records and learned that the Penn Libraries of the University of Pennsylvania created an original record. I then searched, identified and obtained their records. **Appendix 1006C** is a true and correct copy of the bibliographic and MARC records obtained from the Penn Libraries. I personally identified, located, and obtained these records. They are the type of record experts in my field would reasonably rely upon when forming their opinions. I used screen captures to produce this appendix.

33. Pages 1 and 2 of **Appendix 1006C** are the bibliographic record for *Zadrozny* that shows the title, publisher information, format, conference name, other title, subjects, genre, contents, note, and an OCLC number for this record ("80479767").

34. Pages 3 and 4 are the MARC record. The first six digits of Field 008 show this record was created on "950622" (*i.e.*, June 22, 1995). Subfield "a" of Field 040 shows "PU" (*i.e.*, the "University of Pennsylvania Library," according to the

MARC Code List for Organizations) created the original MARC record, and subfield 'c" shows this library entered the original MARC record into OCLC. Field 008 and 040 inform my opinion that the University of Pennsylvania Library cataloged the *Zadrozny* and created an original MARC record for the proceedings on June 22, 1995, making this work discoverable in their library catalog from that date on. Since the MARC record was also entered into OCLC on June 22, 1995, the MARC record also became available to OCLC members for copy cataloging and interlibrary loan activities from that date on. After WorldCat was launched on January 21, 1998, this record from the University of Pennsylvania became discoverable on WorldCat and the record is available at https://search.worldcat.org/title/80479767, the free web portal that supports the general public to explore more than 10,000 libraries worldwide.

35. Field 111 shows the conference name is used in the MARC record as the author, Field 245 shows the title from the title page, Field 264 shows publisher information, field 300 shows physical description, Field 505 shows the content of volumes 1 and 2, and Field 740 shows "CLING94" is included as a variant title to help users discover this proceedings.

36. Subjects are represented in Field 050 by a Library of Congress Classification number of "P98" that represents "General works" on "Computational linguistics. Natural language processing" and in Field 650 by a Library of Congress

subject heading, with "Computational linguistics" as the main topic, followed by a form subheading "Congresses" encoded in subfield "v" to indicate the main topic was treated at a conference. These subject representations support users to conduct subject searches to discover *Zadrozny*.

37. The MARC record from Penn Libraries (**Appendix 1006C**) was created on June 22, 1995, making *Zadrozny* discoverable in the library's catalog from that date on. A customary library practice is to process a newly cataloged work promptly to make it publicly accessible, usually within a week after the cataloging activities are completed. It is therefore my opinion that the physical copy of *Zadrozny*, after processing, would have become publicly accessible at this library by June 29, 1995.

C. Harvard University Library Records

38. Since the Harvard University copy of *Zadrozny* carries a date stamp of "FEB 26 1996," I obtained their library records for this work to determine how users could discover this work back in 1996. **Appendix 1006D** is a true and correct copy of the bibliographic and MARC records for *Zadrozny* that I obtained from the online catalog of Harvard University Library. I personally located, identified, and obtained these records. They are the type of record experts in my field would reasonably rely upon when forming their opinions. I used screen captures to produce this appendix.

39. Pages 1 to 3 are the bibliographic record that shows *Zadrozny* is

available at the Widener Library and its call number if WID-LC P98 .I55 1994X. Page 4 is the MARC record, and the first six digits of Field 008 show the record was created on "960222" (*i.e.*, February 22, 1996). Subfield "a" of Field 040 shows "VYR" created the original MARC record, and subfield "d" shows "HLS" modified the record later. According to the *Directory of OCLC Members* (https://www.oclc.org/en/contacts/libraries.html), "VYR" is the OCLC symbol for the US Air Force Research Laboratory, and "HLS" is the OCLC symbol for the Harvard University. Field 008 and 040 inform my opinion that Harvard University Library used the record created on February 22, 1996, by the US Air Force Research Laboratory library for copy cataloging.

40. Field 111 shows the conference name is used in the MARC record as the author, Field 245 shows the title from the title page, Field 246 shows a variant title, Field 260 shows publisher information, field 300 shows physical description, Field 500 shows this work is "The 15th International Conference on Computational Linguistics."

41. Subjects are represented in Field 090 by a Library of Congress Classification number of "P98" that represents "General works" on "Computational linguistics. Natural language processing" and in Field 650 by a Library of Congress subject heading, with "Computational linguistics" as the main topic, followed by a form subheading "Congresses" encoded in subfield "v" to indicate the main topic

was treated at a conference. These subject representations support users to conduct subject searches to discover *Zadrozny*.

42. The Harvard University MARC record (**Appendix 1006D**) was based on a record created on February 22, 1996, and the library's copy of *Zadrozny* (**Appendix 1006A**) carries a date stamp of "FEB 26 1996." The creation date of the original MARC record by VYR means that by February 22, 1996, the US Air Force Research Laboratory had made *Zadrozny* discoverable in their library catalog. The Harvard MARC record (**Appendix 1006D**) does not explicitly indicate when the copy cataloging work was completed (which is common among most libraries when they do copy cataloging work), but Field 005, a machine-generated field that records the latest date when a MARC record was locally modified, shows "20020606" (*i.e.*, June 6, 2002) was the last date when Harvard University Library modified this record. Record modification can be about fixing typos or updating any data elements in the record. What the "June 6, 2002" indicates is that the Harvard University MARC record for *Zadrozny* would have been in their online catalog no later than June 6, 2002. Considering the date stamp in **Appendix 1006A** and the creation date of the MARC record of VYR that was used as the basis for this library's copy cataloging effort (**Appendix 1006D**), it is my opinion that the copy cataloging work might have been completed in 1996 by the Harvard University Library, and the MARC record would have been in their library catalog no later than June 6, 2002.

This MARC record supports users to discover *Zadrozny* by the conference name, title, and variant title. Users would also have been able to conduct subject searches to discover this proceedings.

D. Internet Archive Records

43. As evidence that the online catalogs of the University of Pennsylvania and Harvard University Library was in operation in 1997 and 1998, I located archived webpages of their online catalogs from Internet Archive. **Appendix 1006E** is a true and correct copy of the webpages of these two online catalogs. I personally located, identified, and obtained these archived webpages. They are the type of material experts in my field would reasonably rely upon when forming their opinions. I used screen captures to produce this appendix.

44. Page 1 of **Appendix 1006E** is a screen capture that shows the webpage of Franklin, the online catalog the University of Pennsylvania on the Web, was archived on January 10, 1998, and available at https://web.archive.org/web/19980110012724/http://www.franklin.library.upenn.edu/. This archived webpage shows Franklin was in operation by January 10, 1998. Page 2 is a screen capture of a webpage that shows users could access Hollis, the online catalog of Harvard University Library. This webpage was archived by Internet Archive on December 10, 1997, and available at https://web.archive.org/web/19971210172035/http://hplus.harvard.edu/alpha/HOL

LIS.html. This archived webpage shows Hollis was in operation by December 10, 1997. **Appendix 1006E** informs my opinion that by December 10, 1997, and January 10, 1998, these two universities had placed their online catalog on the Internet for users to access.

45. Since the MARC record for *Zadrozny* was created by the Penn Libraries of the University of Pennsylvania on June 22, 1995 (**Appendix 1006C**), it is my opinion that users would have been able to search for this work in their library catalog on site in June 1995. After the online catalog became available on the Internet by January 10, 1998, users would have been able to search Franklin on the Internet to discover *Zadrozny*.

46. The Harvard University MARC record (**Appendix 1006D**) would have been discoverable between February 1996 and June 2002. Since their online catalog was available on the Internet by December 10, 1997, users would have been able to use their online catalog on the Internet to discover *Zadrozny* by June 2002, if not earlier.

E. Usage Record

47. Actual usage of a publication is reflected by other publications that make reference to it. I used Google Scholar to review citations. Since *Zadrozny* contains papers presented at the conference, citations were made to those papers individually, not to the entire volume of the proceedings. **Appendix 1006F** is a list

of early citations to papers included in *Zadrozny*. These citations were published between 1995 and 1998, and the earliest citation appeared in 1995, supporting my opinion that *Zadrozny* would have been publicly discoverable from June 22, 1995 on when the original cataloging record was created by the Penn Libraries of the University of Pennsylvania. The physical copy, after processing, would have been publicly accessible at this library by June 29, 1995.

F. Summary of My Opinion on *Zadrozny*

48. The Library of Congress copy of *Zadrozny* (**Ex. 1006**) is authentic because it contains the same content as the copies from Harvard University Library (**Appendix 1006A**) and the Penn Libraries of the University of Pennsylvania (**Appendix 1006B**). The MARC record of the Penn Libraries (**Appendix 1006C**) shows the library created an original record for **Ex. 1006** on June 22, 1995, making this work discoverable in their library catalog from then on. The physical copy, after processing, would have become publicly accessible at this library by June 29, 1995. Citation records show the earliest citation appeared in 1995, supporting my opinion that **Ex. 1006** was publicly discoverable and accessible by June 1995.

IV. Authenticity and Public Availability of *Gibbon* (Ex. 1008)

A. Authentication

49. **Exhibit 1008** is a true and correct copy of portions of *Handbook of Multimodal and Spoken Dialogue Systems: Resources, Terminology and Product*

Evaluation, ("*Gibbon*"), edited by D. Gibbon, I. Martins, and R. K. Moore, and published by Kluwer Academic Publishers, that I received from counsel. This copy was obtained by counsel from Library of Congress. It is presented as **Ex. 1008** in this declaration.

50. Page 1 is the spine that shows the book title, editors, a call number label of "QA 76.9 .N38 H362 2000 Copy 1" and a label of "FT MEADE GenColl" that indicates this book is in the general collection at the Fort Meade storage facility of the Library of Congress. Page 2 is the cover that shows the title, editors, and a notes indicating "CD ROM included" with this book. Page 4 is the half title page. Page 5 is the series title page showing *Gibbon* belongs to "The Kluwer International Series in Engineering and Computer Science." Page 6 is the title page that shows the title, editors, and the publisher. Page 7 is the copyright page that shows *Gibbon* has a "2000" copyright date and Kluwer Academic Publishers holds the copyright. Pages 10 to 14 are the table of contents that shows *Gibbon* includes five chapters, a bibliography, two appendices, a list of abbreviations, an index, and a CD-ROM disclaimer. The table of contents is followed by a list of figures and a list of tables. Page 19 is the Editorial Preface. Pages 20 to 56 are pages of terms in the EAGLET term database, including internal pages [367] to 371, 390 to 391, 424 to 425, 428 to 433, 444 to 445, 448 to 451, 454 to 455, 460 to 461, 464 to 465, 474 to 475, 478 to 479, 486 to 489, and 492 to 493. These terms are selected for their relevance to the

patent in question. Page 57 is the back cover that describes the scope of the book and shows it also includes "[a] fully searchable CD-ROM containing a hypertext version of the book in HTML format for fast look up of specific points, convenient desktop use, and lightweight mobile reference, and [t]he first systematic medium scale compendium of terminology with definitions." The back cover shows "SECS 565" above the book's ISBN of "0-7923-7904-7."

51. **Ex. 1008** is in a condition that creates no suspicion about its authenticity. Specifically, the text on each page of *Gibbon* appears to flow seamlessly, and there are no visible alterations to the pages. Based on my analysis, it is my opinion that **Ex. 1008** is an authentic copy of *Gibbon* and I see no reason to question its authenticity.

B. Library of Congress Records

52. To determine when *Gibbon* became publicly accessible I searched the online catalog of the Library of Congress for records for this book. **Appendix 1008A** is a true and correct copy of the bibliographic and MARC records for this book. I personally identified, located, and obtained these records. They are the type of record experts in my field would reasonably rely upon when forming their opinions.

53. Pages 1 and 2 of **Appendix 1008A** are the bibliographic record for *Gibbon* that shows the title, publisher, ISBN ("0792379047"), related names, subjects, LCCN ("00044780"), and other bibliographic details. The title, publisher,

ISBN, LCCN, and editors match the information presented in **Ex. 1008**, confirming this record represents *Gibbon* held by Library of Congress. The Item Availability area shows the library holds two copies of this book, both are stored offsite and available for user request, and interested users can request access at the Jefferson or Adams Building Reading Rooms.

54. Pages 3 and 4 are the MARC record for *Gibbon*. The first six digits of Field 008 show this record was created on "000607" (*i.e.*, June 7, 2000), and the "s2000" code following the record creation date indicates the book has a single publication date of 2000. Field 955 records local cataloging activities, and it shows the initial MARC record was sent to the Dewey Classification Division on "06-12-00" (*i.e.*, June 12, 2000). The copyright page of **Ex. 1008** shows a Cataloging-in-Publication (CIP) record, and the date stamp shows a CIP receipt date of "AUG 29 2000." These dates inform my opinion that Library of Congress created a CIP record in MARC for *Gibbon* before its publication, and the Cataloging in Publication Program received the book, after it was published, on August 29, 2000. Field 955 shows "CIP ver. … to SL 10-03-00," "to BCCD 10-13-00; copy 2 added … to BCCD 10-30-00." These dates indicate that after record verification, the book was sent to BCCD (Binding and Collections Care) for processing on October 13, 2000, and copy 2 was sent to BCCD on October 30, 2000. Subfield "a" of Field 040 shows "DLC" (*i.e.*, the OCLC symbol for Library of Congress) created the MARC record. Field

005 is a machine-generated field that records the last time the MARC record was modified at Library of Congress, and it shows the last record modification date was "20001030102513" (*i.e.*, October 30, 2000, at 10 a.m., 25 minutes, and 13 seconds). Fields 008, 955, 050 and 040 inform my opinion that Library of Congress cataloged *Gibbon* and created a CIP record on June 7, 2000, before the book was published. The library received a copy of the book after it was published, used it for CIP verification, and copy 1 was sent to the processing unit on October 13, 2000. A customary library practice is to process a newly cataloged book promptly for user access, typically within a week after the cataloging work is completed. My personal experience with the collections at the Library of Congress is that it can take one to three months for a newly cataloged book to be ready for public access. It is therefore my opinion that *Gibbon* became discoverable in the online catalog of the Library of Congress on June 7 , 2000, when the CIP record was created, and the physical copy, after processing, would have become publicly accessible as early as November 13, 2000, and no later than January 13, 2001.

55. Field 010 shows the LCCN, Field 020 shows the ISBN, Field 245 shows the title, subtitle, and statement of responsibility, Field 260 shows the publisher and the copyright date of 2000, Field 300 shows the physical description, Field 440 shows the publisher series and the number of *Gibbon* in this series, and

Fields 700 show the three editors are included in the MARC record to help users discover this book.

56. Subjects are represented in Field 050 by a Library of Congress Classification (LCC) number of "QA76.9.N38" that represents "Natural language processing," and in Field 082 by a Dewey Decimal Classification (DDC) number of "006.3/5" that represents "Natural language processing." Subject is also represented in Fields 650 by twos Library of Congress subject headings, with "Natural language processing (Computer science)" and "Automatic speech recognition" as the main topics. The classification numbers and subject headings support users to conduct subject searches to discover *Gibbon*.

57. This MARC record (**Appendix 1008A)** has made *Gibbon* discoverable at the Library of Congress. The date stamp on **Ex. 1008** shows the book was received by CIP on August 29, 2000, and the MARC record shows that a CIP record was created on June 7, 2000, making *Gibbon* discoverable at the online catalog of the library from that date on. Users would have been able to conduct subject searches by the LCC number, the DDC number, and the Library of Congress subject headings in the MARC record to discover this book. They would also have been able to search for it by the ISBN, LCCN, title, series, and editors.

58. The date stamp on the library copy of *Gibbon* (**Ex. 1008**) shows a receipt date of August 29, 2000, and the MARC record (**Appendix 1008A**) shows

the book was sent to processing on October 13, 2000. It is therefore my opinion that the physical copy, after processing, would have become publicly accessible as early as November 13, 2000, and no later than January 13, 2001.

C. Copyright Registration Record

59. **Appendix 1008B** is a true and correct copy of the copyright registration record for *Gibbon* that I obtained from the public catalog of the US Copyright Office. I personally located, identified, and obtained this record. This is the type of record experts in my field would reasonably rely upon when forming their opinions.

60. I searched the public catalog of the Copyright Office by the title of *Gibbon* to locate the copyright registration record. The title, imprint, series and numbering, and copyright claimant match the information presented in **Ex. 1008**, confirming this record is for *Gibbon*. The "Date of Publication" field shows this book was published on "2000-08-14" (*i.e.*, August 14, 2000), and the "Registration Number / Date" field shows it was registered with the US Copyright Office on "2000-09-19" (*i.e.*, September 19, 2000). These dates align well with the cataloging dates in the MARC record (**Appendix 1008A**). Specifically, these dates inform my opinion that before *Gibbon* was published, Library of Congress created a CIP record for it on June 7, 2000. After the book was published on August 14, 2000, the Cataloging in Publication Program of Library of Congress received a copy on

August 29, 2000, and the copy was sent to processing on October 13, 2000, after record verification was completed.

61. As evidence that the online catalog of the Library of Congress was in operation by 2000, I searched Internet Archive for records. **Appendix 1008C** is a true and correct archived copy of the Library of Congress online catalog. I personally located, identified, and obtained this record. This is the type of record experts in my field would reasonably rely upon when forming their opinions.

62. Page 1 is a screen capture of a Wayback Machine calendar that shows the online catalog (http://catalog.loc.gov/) was archived 5,797 times between "12 Oct 1999" and "2 Apr 2024" and this particular copy was archived on "FEB 29 2000" and available at https://web.archive.org/web/20000229084601/http://catalog.loc.gov/. The archive date means this online catalog would have been publicly accessible on the Internet no later than February 29, 2000. Page 2 is the PDF of this online catalog, and it shows that by February 29, 2000, users could browse the library catalog by subject, names, corporate body, call number, and search by title or serial title. In addition, they could conduct searches by "guided keyword," "command key" or "keyword" in general, and could also set search limits to make the search results more manageable. **Appendix 1008C** informs my opinion the Library of Congress online catalog was in operation by February 2000 on the Internet. Since the library's MARC CIP record for *Gibbon* was completed on June 7, 2000, from that date on, users

would have been able to search the Library of Congress online catalog on the Internet to discover *Gibbon* (**Ex. 1008**).

D. Usage Record

63. Actual usage of a publication is reflected by other publications that make reference to it. I used Google Scholar to review citations to *Gibbon* and compile a list of five citations published in 2001 and 2002 to show early usage (**Appendix 1008D**). The earliest citation appeared in May 2001.

E. Summary of My Opinion on *Gibbon*

64. **Ex. 1008** is authentic because the text in *Gibbon* contains no signs of alterations, the copy raises no suspicion about its authenticity, and the book is in the custody of a national library. The MARC record (**Appendix 1008A**) shows Library of Congress created a pre-publication Cataloging-in-Publication record for this book on June 7, 2000, making it discoverable at the library from that date on. After the book was published on August 14, 2000, as shown by the copyright registration record (**Appendix 1008B**), the Cataloging in Publication Program received a copy on August 29, 2000, completed record verification, and sent the copy to processing on October 13, 2000. It is therefore my opinion that *Gibbon* became discoverable on June 7, 2000, and the physical copy, after processing, would have become publicly accessible no later than January 13, 2001. Citation history of this book shows the earliest citation to *Gibbon* appeared in May 2001 (**Appendix 1008D**).

V. Authenticity and Public Availability of *Sutcliffe* (Ex. 1016)

A. Authentication

65. **Exhibit 1016** is a true and correct copy of portions of *Industrial Parsing of Software Manual, ("Sutcliffe"),* editors Richard F.E. Sutcliffe, Heinz-Detlev Koch and Annette McElligott, published by Rodopi in 1996, that I received from counsel. This copy was obtained by counsel from the Memorial Library of the University of Wisconsin-Madison. It is presented as **Ex. 1016** in this declaration.

66. Page 1 is the spine of a book that shows the title, editors, and a call number label of "PE 1074.5 I48 1996." Page 2 is the title page that shows the title, editors, and publisher. Page 3 is the copyright page that shows this book has a "1996" copyright date and Rodopi B.V. holds the copyright. It also shows the ISBNs and a library stamp. Pages 4 to 24 (internal pages [25] to 45) are the full text of Chapter 3 "Comparative Evaluation of Grammatical Annotation Models" by Eric Steven Atwell. This chapter has a total of 21 pages, including four tables and 20 references.

67. **Ex. 1016** is in a condition that creates no suspicion about its authenticity. Specifically, the text on each page of *Sutcliffe* appears to flow seamlessly from one page to the next, and there are no visible alterations to the pages. In addition, this book is in the custody of an academic library. Based on my analysis, it is my opinion that **Ex. 1016** is an authentic copy of *Sutcliffe* and I see no reason to question its authenticity.

B. University of Wisconsin-Madison Library Records

68. To determine when *Sutcliffe* became publicly accessible I searched the online catalog of the University of Wisconsin-Madison Library System for records for this book. **Appendix 1016A** is a true and correct copy of the bibliographic and MARC records for this book. I personally identified, located, and obtained these records. They are the type of record experts in my field would reasonably rely upon when forming their opinions. I used screen captures to produce this appendix.

69. Pages 1 and 2 of **Appendix 1016A** are the bibliographic record for *Sutcliffe* that shows the title, call number, editors, format, language, publisher, physical description, ISBNs, and OCLC number ("36467363"). The title, call number ("PE1074.5 I48 1996"), editors, ISBN ("9042001143") and publisher information match the information presented in **Ex. 1016**, confirming this record represents *Sutcliffe*.

70. Pages 3 and 4 are the MARC record for *Sutcliffe*. The first six digits of Field 008 show this record was created on "970228" (*i.e.*, February 28, 1997). Subfield "a" of Field 040 shows "NYP (the OCLC symbol for the New York Public Library) created the original MARC record, and subfield "d" shows "GZM" (the OCLC symbol for the University of Wisconsin-Madison libraries) modified the original record later. Fields 008 and 040 inform my opinion that New York Public Library created a MARC record for *Sutcliffe* on February 28, 1997, and the

University of Wisconsin-Madison Library System used the original record for copy cataloging later.

71. Field 020 shows the ISBN, Field 245 shows the title and statement of responsibility, Field 260 shows the publisher and the publication date, Field 300 shows the physical description, and Fields 700 show the three editors are included in the MARC record to help users discover this book.

72. Subjects are represented in Field 090 by a Library of Congress Classification (LCC) number of "PE1074.5" that represents "Language data processing," and in Field 650 by a Library of Congress subject heading "Linguistic String Parser (Computer grammar)." The classification number and subject heading support users to conduct subject searches to discover *Sutcliffe*.

73. This MARC record (**Appendix 1016A)** has made *Sutcliffe* discoverable at the University of Wisconsin-Madison Library System. **Ex. 1016** does not carry a date stamp, because most libraries, except for national libraries, do not date stamp books they acquire. The MARC record (**Appendix 1016A**) shows the library used a New York Public Library record as the basis for copy cataloging, but it does not indicate when the copy cataloging record was completed. Nevertheless, **Ex. 1016** and **Appendix 1016A** show *Sutcliffe* was available at the University of Wisconsin-Madison Library System, and interested users could use the ISBN, title, and editors to discover this book. They could also conduct subject searches to locate this book.

C. University of California Berkely Libraries Records

74. To determine when *Sutcliffe* became publicly accessible at libraries, I searched WorldCat for records, and a record at https://search.worldcat.org/title/36467363 showed that the University of California Berkeley Libraries was another holding library of this work, so I obtained this library's records for analysis. **Appendix 1016B** is a true and correct copy of the bibliographic and MARC records for *Sutcliffe* that I obtained from the online catalog of the University of California Berkeley Libraries. I personally located, identified, and obtained this record. This is the type of record experts in my field would reasonably rely upon when forming their opinions. I used screen captures to produce this appendix.

75. Pages 1 and 2 of **Appendix 1016B** are the bibliographic record for *Sutcliffe*. Page 1 shows the book title, editors, and publication year, and indicates this book is available at Doe Library with a call number of PE1074.5 .I48 1996. Page 2 shows bibliographic details, including the title, format, creation date, publisher, series, notes, subjects, ISBNs.

76. Pages 3 and 4 are the MARC record. The first six digits of Field 008 show the record was created on "970228" (*i.e.*, February 28, 1997). Subfield "a" of Field 040 shows "NYP" (the OCLC symbol for the New York Public Library System) created the original record, and subfield "d" shows "CUY" (the OCLC

symbol for the University of California Berkeley Libraries) used the original MARC record for copy cataloging and modified the original record. The "Addate" data element in Field 968 on page 4 shows *Sutcliffe* was added to the library collection on "970409" (*i.e.*, April 9, 1997), and the "CatDate" element in Field 996 shows the local cataloging record was created on "04-09-97" (*i.e.*, April 9, 1997). Fields 008, 040, 968 and 996 inform my opinion that the University of California Berkeley Libraries used the original MARC record created by the New York Public Library System on February 28, 1997, as the basis for their copy cataloging, and completed their copy catalog record on April 9, 1997, making *Sutcliffe* discoverable from that date on in their online catalog.

77. Since the University of Wisconsin MARC record (**Appendix 1016A**) and the University of California Berkeley MARC record (**Appendix 1016B**) were based on the same New York Public Library MARC record, these two MARC records have many data in common, including the record creation date of the original MARC record in Field 008, the ISBN for the bound volume (Field 020), OCLC number (Field 035), title (Field 245), publisher (Field 260), physical description (Field 300), bibliographical reference and index note (Field 504), editors (Fields 700), call number (Field 090), and subject heading (Field 650). One difference is the University of Wisconsin MARC record does not include the series information.

78. **Appendix 1016B** informs my opinion that *Sutcliffe* was added to the University of California Berkeley Library on April 9, 1997, and the copy cataloging record was completed on the same day, making *Sutcliffe* discoverable. Users would have been able to use the ISBN, title, editors, classification number, and subject heading to discover this work. As discussed above, a common library practice is to process a newly cataloged book promptly for public access, usually within a week after the cataloging work is completed. It is therefore my opinion that the physical copy of *Sutcliffe*, after processing, would have become publicly accessible at this library no later than April 16, 1997.

79. As evidence that the online catalog of the University of California Berkeley Library was in operation by 1997, I searched Internet Archive for records. **Appendix 1016C** is a true and correct archived copy of the "Library Online Catalog Systems" available at the University of California Berkeley Library and the rest of the University of California system. I personally located, identified, and obtained this record. This is the type of record experts in my field would reasonably rely upon when forming their opinions.

80. Page 1 is a screen capture of a Wayback Machine calendar that shows this webpage of "Library Online Catalog Systems" (http://www.lib.berkeley.edu/Catalogs/) was archived 746 times between "7 Jun 1997" and "22 Aug 2023" and this particular copy was archived on "JUN 07 1997"

and available at https://web.archive.org/web/19970607084855/http://www.lib.berkeley.edu/Catalogs/ . The archive date means that by June 7, 1997, this online catalog systems for the University of California Berkeley Library users and other students of the university system would have been publicly accessible. Page 2 is the PDF of this archived webpage, and it shows that by June 7, 1997, GLADIS, Melvyl and other catalogs were available to help users explore collections at the University of California systems. **Appendix 1016C** informs my opinion the online catalog systems of the University of California Berkeley Library were in operation by June 1997 on the Internet. Since the library's MARC record for *Sutcliffe* was completed on April 9, 1997, by June 7, 1997, users would have been able to search the library catalog on the Internet to discover *Sutcliffe* (**Ex. 1016**).

D. Usage Record

81. Actual usage of a publication is reflected by other publications that make reference to it. I used Google Scholar to review citations to *Sutcliffe* and compile a list of five citations published from 1997 to 2001 to show usage (**Appendix 1016D**). The earliest citation appeared in September 1997.

E. Summary of My Opinion on *Sutcliffe*

82. **Ex. 1016** is authentic because the text in *Sutcliffe* contains no signs of alterations, the copy raises no suspicion about its authenticity, and the book is in the

custody of an academic library. The University of Wisconsin MARC record (**Appendix 1016A**) shows users could discover *Sutcliffe* by ISBN, title, editors, classification number and subject heading. The University of California Berkeley MARC record (**Appendix 1016B**) shows *Sutcliffe* was added to the library collection on April 9, 1997, and the library's copy cataloging record was completed on April 9, 1997, making *Sutcliffe* discoverable from that date on in their online catalog. Users would have been able to search for *Sutcliffe* by ISBNs, title, editors, classification number, and subject heading. The physical copy of *Sutcliffe*, after processing, would have become publicly accessible by April 16, 1997. Citation history shows the earliest citation to *Sutcliffe* appeared in September 1997 (**Appendix 1016D**).

VI. Conclusion

83. I hereby declare that all statements made herein on my own knowledge are true and that all statements made on information and belief are believed to be true, and further, that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under 18 U.S.C. § 1001.

Date: 4-2-2024 Executed: Ingrid Hsieh-Yee

Ingrid Hsieh-Yee, Ph.D.
Annapolis, Maryland

Appendix 1006A

COLING 94

The 15th International Conference on Computational Linguistics

PROCEEDINGS

Vol. II

August 5 – 9, 1994

Kyoto, Japan

Google Exhibit 1020
Page 0045

WID-LC
P
98
.I55
1994x
Vol. 2

HARVARD UNIVERSITY
LIBRARY
FEB 26 1996
25773

Table of Contents (Vol. II)

Semantics

Word Sense Acquisition for Multilingual Text Interpretation ---------- 665
P. S. Jacobs (GE Research and Development Center)
A System of Verbal Semantic Attributes Focused on the Syntactic Correspondence between Japanese and English ---------- 672
H. Nakaiwa, A. Yokoo, S. Ikehara (NTT Communication Science Labs.)
Semantics of Complex Sentences in Japanese ---------- 679
H. Nakagawa, S. Nishizawa (Yokohama National University)
Content Characterization Using Word Shape Tokens ---------- 686
P. Sibun, D. S.Farrar (Fuji Xerox Palo Alto Lab.)
The Nature of Near-synonymic Relations ---------- 691
C. DiMarco (University of Waterloo)
Virtual Polysemy ---------- 696
A. Sanfilippo, K. Benkerimi, D. Dwehus (Sharp Labs. of Europe)
Building a Windows-based Bilingual Functional Semantic Processor ---------- 701
J. J. Webster (City Polytechnic of Hong Kong)
On the Proper Role of Coercion in Semantic Typing ---------- 706
J. Pustejovsky, P. Bouillon (Brandeis Univ./Univ. of Geneva)
Word Sense Ambiguation:Clustering Related Senses ---------- 712
W. B. Dolan (Microsoft Corp.)
A Best-match Algorithm for Broad-coverage Example-based Disambiguation ---------- 717
N. Uramoto (IBM Japan Ltd.)
Drawing Pictures with Natural Language and Direct Manipulation ---------- 722
M. Hiyoshi, H. Shimazu (NEC Corp.)
Verbal Case Frame Acquisition from a Bilingual Corpus:Gradual Knowledge Acquisition --- 727
H. Tanaka (NHK Science & Technical Research Labs.)
An Empirical Study on the Generation of Zero Anaphors in Chinese ---------- 732
C.-L. Yeh, C. Mellish (University of Edinburgh)
Representing Information Need with Semantic Relations ---------- 737
A. S. Chakravarthy (MIT Media Lab.)
Generalizing Automatically Generated Selectional Patterns ---------- 742
R. Grishman, J. Sterling (New York University)
Incremental Interpretation: Applications, Theory, and Relationship to Dynamics Semantics 748
D. Milward, R. Cooper (University of Edinburgh)
Word Sense Disambiguation and Text Segmentation Based on Lexical Cohesion ---------- 755
M. Okumura, T. Honda (Japan Advanced Institute of Science and Technology)
Automatic Recognition of Verbal Polysemy ---------- 762
F. Fukumoto, J. Tsujii (UMIST)
An Experiment on Learning Appropriate Selectional Restrictions from a Parsed Corpus --- 769
F. R. Framis (Universitat Politècnica de Catalunya)

Google Exhibit 1020
Page 0047

A Discrete Model of Degree Concept in Natural Language ---------- 775
S. Kamei, K. Muraki (NEC Corp.)
Algorithm for Automatic Interpretation of Noun Sequences ----------782
L. Vanderwende (Microsoft Corp.)
A Treatment of Functional Definite Descriptions ----------789
H. Wada (Intelligent Text Processing, Inc.)
Bottom-up Earley Deduction ----------796
G. Erbach (University of Saarland)
The Merged Upper Model:A Linguistic Ontology for German and English ----------803
R. Henschel, J. Bateman (GMD/IIPI)

Syntax

Modeling Dialogue by Functional Subcategorization ----------813
J.R. Zubizarreta Aizpuru, C. Jones (U.P.V.–E.H.U./Univ. of Aberdeen)
Dutch Cross Serial Dependencies in HPSG ----------818
G. Rentier (Tilburg University)
HPSG Lexicon without Lexical Rules ----------823
K. Oliva (Universiy of Saarland)
A Lexicon of Distributed Noun Representations Constructed by Taxonomic Traversal ------- 827
R. Sutcliffe, D. O'Sullivan, F. Meharg (University of Limerick)
Multi-modal Definite Clause Grammar ----------832
H. Shimazu, S. Arita, Y. Takashima (NEC Corp.)
Hypothesis Selection in Grammar Acquisition ----------837
M. Kiyono, J. Tsujii (UMIST)
Achieving Flexibility in Unification Formalisms ----------842
L. Strömbäck (Linköping University)
TWP:How to Assist English Production on Japanese Word Processor ----------847
K. Muraki, S. Akamine, K. Satoh, S. Ando (NEC Corp.)
Long-distance Dependencies and Applicative Universal Grammar ----------853
S. Shaumyan, F. Segond (Yale University/Rank Xerox Res. Centre)
A Reestimation Algorithm for Probabilistic Recursive Transition Network ----------859
Y. S. Han, K.-S. Choi (Korea Advanced Institute of Science and Technology)
Analysis of Japanese Compound Nouns Using Collocational Information ----------865
Y. Kobayashi, T. Tokunaga, H. Tanaka (Tokyo Institute of Technology)
Free-ordered CUG on Chemical Abstract Machine ----------870
S. Tojo (Mitsubishi Research Institute, Inc.)
Computing FIRST and FOLLOW Functions for Feature-theoretic Grammars ----------875
A. Trujillo (University of Cambridge)
Focus on "Only" and "Not" ----------881
A. Ramsay (University College Dublin)
Structure Sharing Problem and its Solution in Graph Unification ----------886
K. Kogure (NTT Basic Research Labs.)
TDL-A Type Description Language for Constraint-based Grammars ----------893
H.-U. Krieger, U. Schäfer (German Research Center for Artificial Intelligence)
On the Portability of Complex Constraint-based Grammars ----------900
C.J. Rupp, R. Johnson (IDSIA)

Google Exhibit 1020
Page 0048

A Grammar Based Approach to a Grammar Checking of Free Word Order Languages ---- 906
V. Kubon, M. Plátek (Charles University)
Table-driven Neural Syntactic Analysis of Spoken Korea ---------- 911
W. Lee, G. Lee, J.-H. Lee (Phonag Institute of Science and Technology)
Universal Guides and Finiteness and Symmetry of Grammar Processing Algorithms ------ 916
M. Martinovic (New York University)
XTAG System-A Wide Coverage Grammar for English ---------- 922
C. Doran, D. Egedi, B. A. Hockey, B.Srinivas, M. Zaidel (University of Pennsylvania)
Weakly Restricted Stochastic Grammars ---------- 929
R. O. D. Akker, H. ter Doest (University of Twente)
Non-constituent Coordination:Theory and Practice ---------- 935
D. Milward (University of Edinburgh)
Hypothesis Scoring over Theta Grids Information in Parsing Chinese Sentences with Serial Verb Constructions ---------- 942
K. H.C. Lin, V.-W. Soo (National Tsing-Hua University)
An Efficient Syntactic Tagging Tool for Corpora ---------- 949
M. Zhou, C. Huang (Tsinghua University)
The Correct and Efficient Implementation of Appropriateness Specifications for Typed Feature Structures ----------956
D. Gerdemann, P. J. King (Universität Tübingen)
A Classification Method for Japanese Signs Using Manual Motion Descriptions ----------961
H. Adachi, K. Kamata (Utsunomiya University)

Speech & Phonology

Machine-readable Dictionaries in Text-to-speech Systems ----------971
J. L. Klavans, E. Tzoukermann (Columbia University)
Issues in Text-to-speech for French ----------976
E. Tzoukermann (AT&T Bell Labs.)
CHATR: A Generic Speech Synthesis System ----------983
A. W. Black, P. Taylor (ATR Interpreting Telecommunications Labs.)
Pause as a Phrase Demarcator for Speech and Language Processing ----------987
J. Hosaka, M. Seligman, H. Singer (ATR Interpreting Telephony Research Labs.)
The Parsody System: Automatic Prediction of Prosodic Boundaries for Text-to-speech --- 992
S. Minnis (BT Labs.)
Anytime Algorithms for Speech Parsing? ----------997
G. Görz, M. Kesseler (University of Erlangen-Nürnberg)
Towards a Proper Linguistic and Computational Treatment of Scrambling: An Analysis of Japanese ----------1002
S. Fong (NEC Research Institute, Inc.)
Phonological Derivation in Optimality Theory ----------1007
M. T. Ellison (University of Edinburgh)
A Grammar and a Parser for Spontaneous Speech ----------1014
M. Nakano, A. Shimazu, K. Kogure (NTT Basic Research Labs.)
Catching the Cheshire Cat ----------1021
C. Johansson (Lund University)

Google Exhibit 1020
Page 0049

Information Retrieval & Extraction

A Dutch to SQL Database Interface Using Generalized Quantifier Theory ---------- 1029
D. Speelman, G. Adriaens (University of Leuven)
A Methodology for Automatic Term Recognition ---------- 1034
S. Ananiadou (Manchester Metropolitan University)
Knowledge Extraction from Texts: A Method for Extracting Predicate-argument Structures from Texts ---------- 1039
F. Pugeault, P. Saint-Dizier, M.-G. Monteil (Université Paul Sabatier/EDF, D.E.R.)
Thesaurus-based Efficient Example Retrieval by Generating Retrieval Queries from Similarities ---------- 1044
T. Utsuro, K. Uchimoto, M. Matsumoto, M. Nagao
(Nara Institute of Science and Technology/Kyoto University)
Tools for Extracting and Structuring Knowledge from Texts ---------- 1049
A. Ogonowski, M. L. Herviou, E. Dauphin (GSI-ERLI/AEROSPATIALE CCR/EDF)
N-Gram Cluster Identification During Empirical Knowledge Representation Generation -- 1054
R. Collier (University of Sheffield)
Document Classification by Machine: Theory and Practice ---------- 1059
L. Guthrie, E. Walker (New Mexico State University)
Pattern Matching in the Textract Information Extraction System ---------- 1064
T. Kitani, Y. Eriguchi, M. Hara (Carnegie Mellon University)
Recognizinig Text Genres with Simple Metrics Using Discriminant Analysis ---------- 1071
J. Karlgren, D. Cutting (Swedish Institute of Computer Science/Apple Computer)
Bilingual Text Matching Using Bilingual Dictionary and Statistics ---------- 1076
T. Utsuro, H. Ikeda, M. Yamane, Y. Matsumoto, M. Nagao
(Nara Institute of Science and Technology/Kyoto University)
Towards a More User-friendly Correction ---------- 1083
D. Genthial, J. Courtin, J. Ménézo (LGI-IMAG)
Reverse Queries in DATR ---------- 1089
H. Langer (University of Osnabrück)
K-vec: A New Approach for Aligning Parallel Texts ---------- 1096
P. Fung, K. W. Church (Columbia University/AT&T Bell Labs.)

Discourse & Pragmatics

A Formal Representation of the Thematic-rhematic Structure of Sentences Based of a Typed A-calculus ---------- 1105
Y. Uetake (Tokyo University of Mercantile Marine)
Customizing and Evaluating a Multilingual Discourse Module ---------- 1109
C. Aone (Systems Research and Applications Corp. [SRA])
Consequence Relations in DRT ---------- 1114
S. Akama, Y. Nakayama (Teikyo Univ. of Technology/Nihon Unisys Ltd.)
Collaboration on Reference to Objects that are not Mutually Known ---------- 1118
P. G. Edmonds (University of Toronto)
Automatic Detection of Discourse Structure by Checking Surface Information in Sentences 1123
S. Kurohashi, M. Nagao (Kyoto University)

Google Exhibit 1020
Page 0050

Extending DRT with a Focusing Mechanism for Pronominal Anaphora and Ellipsis Resolution ---------- 1128
J. Abraços, J. G. Lopes (CRIA/UNINOVA Faculdade de Ciências e Tecnologia)
Reference Resolution Using Semantic Patterns in Japanese Newspaper Articles ---------- 1133
T. Wakao (University of Sheffield)
Exploiting Reference Interaction in Resolving Temporal Reference ---------- 1138
K. Dohsaka (NTT Basic Research Labs.)
A Grammatico-statistical Approach to Discourse Partitioning ---------- 1145
T. Nomoto, Y. Nitta (Hitachi Ltd., ARL)
Centering in Japanese: A Step Towards Better Interpretation of Pronouns and Zero-pronouns ---------- 1151
S. Takada, N. Doi (Keio University)
Robust Method of Pronoun Resolution Using Full-text Information ---------- 1157
T. Nasukawa (IBM Research)
Towards a Dynamic Theory of Belief-sharing in Cooperative Dialogues ---------- 1164
H. Komatsu, N. Ogata, A. Ishikawa (Toin Corp./Univ. of Tsukuba/Sophia Univ.)
An Integrated Model for Anaphora Resolution ---------- 1170
R. Mitkov (Institute of Mathematics)
Breaking Down Rhetorical Relations for the Purpose of Analysing Discourse Structures -- 1177
J. Fukumoto, J. Tsujii (UMIST)
Presupposition & VP-Ellipsis ---------- 1184
J. Bos (Universität des Saarlandes)
Communicating with Multiple Agents ---------- 1191
E. A. Hinkelman, S. P. Spackman (DFKI)
A Rule-based Approach to Prepositional Phrase Attachment Disambiguation ---------- 1198
E. Brill, P. Resnik (Massachusetts Institute of Technology/Sun Microsystems Labs.)
Discourse and Deliberation: Testing a Collaborative Strategy ---------- 1205
M. A. Walker (Mitsubishi Electric Research Labs.)
A Bayesian Approach for User Modeling in Dialogue Systems ---------- 1212
T. Akiba, H. Tanaka (Tokyo Institute of Technology)

Google Exhibit 1020
Page 0051

Reserve Paper

World Class Discovery for Postprocessing Chinese Handwriting Recognition ---------- 1221
C.-H. Chang (Industrial Technology Research Institute)
A Simple Transformation for Offline-Parsable Grammars and Its Termination Properties -- 1226
M. Dymetman (Rank Xerox Research Center)
The Evaluation of Machine-Tractable Dictionaries ---------- 1231
C.-M. Guo, C.-N. Huang, J.-P. Gong, J. Li (Tsinghua University)
Discontinuity and the Lambek Calculus ---------- 1235
M. Hepple (University of Sheffield)
Lexical Functions and Machine Translation ---------- 1240
D. Heylen, K.G. Maxwell, M. Verhagen (OTS)
Chinese Segmentation Disambiguation ---------- 1245
W. Jin (New Mexico State University)
Typed Feature Structures as Descriptions ---------- 1250
P.J. King (Eberhard-Karls-Universität)
Portuguese Analysis with Tree Adjoining Grammars ---------- 1255
K.C. Kipper, V.L.S. de Lima (PUCRS)
Incremental Construction of a Lexical Transducer for Korean ---------- 1262
H.-C. Kwon, L. Karttunen (Pusan National University)
Dynamic Logic with Possible World ---------- 1267
R. Lu (Shanghai Jiao Tong University)
Humor-based Applications ---------- 1270
G. Prószéky, M. Pál, L. Tihanyi (Morphologic/OPKM Comp. Centre/Inst. for Linguistics)
Blending Segmentation with Tagging in Chinese Language Corpus Processing ---------- 1274
Z. Qiang, Y. Shiwen (Peking University)
Syllable-based Phonetic Transcription by Maximum Likelihood Methods ---------- 1279
R.A. Sharman (IBM UK Labs., Ltd.)
A Knowledge Acquisition and Management System for Morphological Dictionaries --------- 1284
P. Ten Hacken, S. Bopp, M. Domening, D. Holz, A. Hsiung, S. Pedrazzini
(Universitat Basel/Vrije Univ. Amsterdam/IDSIA)
NL Understanding with a Grammar of Constructions ---------- 1289
W. Zadrozny, M. Szummer, S. Jarecki, D.E. Johnson, L. Morgenstern (IBM)

Google Exhibit 1020
Page 0052

NL Understanding with a Grammar of Constructions

Wlodek Zadrozny and Marcin Szummer and Stanislaw Jarecki
and David E. Johnson and Leora Morgenstern *
IBM Research Division
T.J.Watson Research Lab
Yorktown Heights, NY 10598 USA

Abstract

We present an approach to natural language understanding based on a *computable grammar of constructions*. A *construction* consists of a set of features of form and a description of meaning in a context. A grammar is a set of constructions. This kind of grammar is the key element of MINCAL, an implemented natural language speech-enabled interface to an on-line calendar system. The architecture has two key aspects: (a) the use of constructions, integrating descriptions of form, meaning and context into one whole; and (b) the separation of *domain knowledge* (about calendars) from *application knowledge* (about the particular on-line calendar).

1 Introduction: an overview of the system

We present an approach to natural language understanding based on a *computable grammar of constructions*. A *construction* consists of a set of features of form and a description of meaning in a context. A grammar is a set of constructions. This kind of grammar is the key element of MINCAL, an implemented natural language speech-enabled interface to an on-line calendar system.

The system consists of a NL grammar, a parser, an on-line calendar, a domain knowledge base (about dates, times and meetings), an application knowledge base (about the calendar), a speech recognizer, a speech generator.

In this paper we describe two key aspects of the system architecture: (a) the use of constructions, where instead of separating NL processing into the phases of syntax, semantics and pragmatics, we integrate descriptions of form, meaning and context into one whole, and use a parser that takes into account all this information (see [10] for details); (b) the separation of the *domain knowledge* (about calendars) and the *application knowledge* (about the particular on-line calendar).

*M. Szummer and S. Jarecki are also from MIT.

The dialogs

The system allows users to engage in dialogs like:

— *Schedule a meeting with Bob!*
— *At what time and date?*
— *On August 30th.*
— *At what time?*
— *At 8.*
— *Morning or afternoon?*
— *In the evening.*

The parser recognizes *Schedule a meeting with Bob* as an instance of *sent(imp)*, the imperative construction consisting of a verb and an NP, here *np(event)*. The context is used to prevent another reading in which *with Bob* modifies *schedule*, as in *Dance a tango with Bob!*. That is, a contextual rule is used which says that for calendar applications, people do not modify actions or places. Context also plays an important role in understanding answers, e.g. *At 8.* This is understood as a time expression (and not place or rate or something else) only because of the context.

The parameters of a meeting can be given in many ways, e.g. synonyms or different constructions can be used, users can include as many parameters in a sentence as they wish, and the parameters can be given in any order. As a result there are about 10,000 ways of scheduling meetings (with a given set of parameters).

How are the dialogs understood

With respect to parsing, grammars of constructions can be parsed like "standard" grammars, except that the set of features is richer. Given a string (representing a sentence, a fragment of a discourse or a paragraph), the parser assigns it a construction. From this viewpoint, the situation is similar to "regular" parsing, and the possible algorithms are similar. We have implemented a prototype chart parser for construction grammars, discussed further in Section 3.

But, clearly, having understood the sentence as a linguistic entity in isolation is not the ultimate goal. Here the message of an utterance must be understood in the context of an intended action. This is done in two steps. First, the system determines the intended

Google Exhibit 1020
Page 0053

action and its parameters, using domain knowledge (meetings+time+places). Second, once all the parameters have been extracted from the dialog, the system executes the action. To do this, the program uses application-specific knowledge to translate the action and its parameters into a form that can be executed by the application (Xdiary).

2 Constructions as data structures

A construction is given by the matrix:

$$\begin{bmatrix} \mathbf{N} & : \textit{name_of_construction} \\ \begin{bmatrix} \mathbf{C} & : & \textit{context} \\ \mathbf{V} & : & \textit{structure} \\ \mathbf{M} & : & \textit{message} \end{bmatrix} \end{bmatrix}$$

The *vehicle* **V** consists of formulas describing presence (or perhaps absence) of certain *taxemes*, or features of form, within the structure of the construction. Such a structure is given by a list of subconstructions and the way they have been put together (in all our examples this is concatenation, but there are other possibilities, e.g. wrapping). The *context*, **C** , consists of a set of semantic and pragmatic constraints limiting the application of the construction. It can be viewed as a set of *preconditions* that must be satisfied in order for a construction to be used in parsing. The *message*, **M** , describes the meaning of the construction, via a set of syntactic, semantic and pragmatic constraints.

To make this concrete, let us consider a few examples. We begin with a simple "command construction" consisting of an action verb followed by its argument.

$$\begin{bmatrix} \mathbf{N} & : \textit{sent}(\textit{cmnd}, \textit{v.np}) \\ \begin{bmatrix} \mathbf{C} & : [< \textit{hr attends} >= \textit{sr}] \\ \mathbf{V} & : \begin{bmatrix} \textit{struc} = (V.NP) \\ < V\ \textit{cons_n} >= \textit{verb} \\ < V\ \mathbf{M}\ \ \textit{v_type} >= \textit{action_verb} \\ < NP\ \textit{cons_n} >= \textit{np} \end{bmatrix} \\ \mathbf{M} & : \begin{bmatrix} \textit{sem_cat} = \textit{command} \\ \textit{a_type} =< V\ \mathbf{M}\ \ \textit{sem_type} > \\ \textit{a_obj} =< NP\ \mathbf{M}\ \ \textit{sem_type} > \\ \textit{agent} = \textit{hr} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

The context of the construction describes all situations in which the the hearer *hr* (human or machine) is paying attention to the speaker *sr* (a "ready" state). The feature *struc* is a list of variables and/or words/tokens; it is used to describe the structure of a construction, and its role is similar to a rule in a generative grammar. (We will write names of variables in capital letters, e.g. *NP*, inside matrices of constructions). The attribute *cons_n* gives the name of a construction that could be assigned to a string. We use it here to say that the form of the construction can be described as a concatenation of two strings, of which one is a verb (construction) and the other an np (construction). Furthermore, the verb type $< V\ \mathbf{M}\ \ v_type >$ is "action_verb". (The expression $< V\ \mathbf{M}\ \ v_type >$ should be read "the *v_type* of the message of *V*").

The message **M** describes the meaning of the construction as that of a command in which the type of action is described by the meaning of the verb, and the object of the action is given by the meaning of the noun phrase. The attribute *sem_type* stands for the "semantic type" and we identify it currently with the word sense. Thus "erase the file" is understood as a command to *delete* the file, if $< \textit{erase}\ \mathbf{M}\ \ \textit{sem_type} >= \textit{delete}$, but "erase the picture" might refer to the type of action associated with *rub_out*. In both cases the hearer *hr* is supposed to be the agent of the action.

Constructions: from words to discourse

Words, phrases, and fragments of discourse can be analyzed as constructions. We view languages as collections of constructions which range from words to discourse. We claim that the same representation scheme can be used for all constructions.

The examples we are going to present have been developed with a specific purpose in mind, namely for scheduling calendar events. In other papers ([10] and [6]), we have presented examples showing that we can give a good descriptions of non-standard constructions. However, in either case descriptions of meanings and contexts are general, and hence applicable to other tasks.

We now turn our attention to words. The verb "cancel" can be represented as follows:

$$\begin{bmatrix} \mathbf{N} & : \textit{verb}(\textit{cancel}) \\ \begin{bmatrix} \mathbf{C} & : & \begin{bmatrix} \textit{lang_code} & = & \textit{english} \\ \textit{lang_channel} & = & \textit{text} \end{bmatrix} \\ \mathbf{V} & : & \textit{struc} = (\textit{cancel}) \\ \mathbf{M} & : & \begin{bmatrix} \textit{cat} & = & \textit{verb} \\ \textit{sem_type} & = & \textit{delete} \\ \textit{v_type} & = & \textit{action_verb} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Notice that even simple words require context to be (properly) interpreted. In **C** we say that English text is expected (but in other cases it could also be French text, or French speech, etc.). Some aspects of context do not have to be explicitly specified and can be replaced by defaults.

Although the vehicle and the message are both very simple in this example, the simplicity of the message is a result of deliberate simplification. We have restricted it to the specification of the semantic type, identified with one sense of the word, and to describing the verb type of "cancel" as a verb of action. Notice that the other sense of "cancel" – "offset, balance out" – would appear in another entry.

Of course, in reality, the lexical meaning of any word is a much more complicated matter [1]. For instance, in our lexicon the messages of words may contain many of the attributes that appear in the

Google Exhibit 1020
Page 0054

explanatory combinatorial dictionary of Melcuk [7].

Discourse constructions: To illustrate discourse constructions, we consider the following dialog:

Have you arranged the room yet?
No, but I'll do it right away.

We view the pattern of the answer *no.but.S* as a discourse construction. It can represented by the following array of features:

$$\begin{bmatrix} \mathbf{N} : sent(assrt, no.but.S) \\ \begin{bmatrix} \mathbf{C} : [< p_utter\ cons_n >= sent(ques, *)] \\ \mathbf{V} : \begin{bmatrix} struc = (no.but.S) \\ < S\ cons_n >= sent(assrt, *) \end{bmatrix} \\ \mathbf{M} : \begin{bmatrix} < p_sent\ truth_value >= 0 \\ < S\ \mathbf{M} > \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

As we can see, the construction applies only in the context of a previously asked question, and its message says that the answer to the question is negative, after which it elaborates the answer with a sentence *S*.

3 System Architecture

The parts

MINCAL consists of a NL grammar, a parser, a domain knowledge base (about dates, times and meetings), an on-line calendar (Xdiary), an application knowledge base (about Xdiary), a continuous speech recognizer (IBM, ICSS), a speech generator (Speech Plus, Text to Speech Converter), and the interfaces.

At present, the grammar consists of a few hundred lexical constructions, and about 120 "productions", i.e. constructions describing combinations of other constructions. [1] It covers the basic forms of assertive sentences, but it emphasizes commands. Thus a command can, for example, be given either by v.np (also with "please", or "kindly"), or by an assertive sentence recognized as an indirect speech act ("I'd like to ...", "Leora wants you to ...", etc.). The next large group of constructions covers PPs, with particular emphasis on time and places. Finally, it covers a few discourse constructions, since it is important to deal with sentence fragments in dialogs, e.g. understanding "evening" as "in the evening", when it is an answer to the question "when?".

The interaction of the modules

The calendar and the application knowledge base: Xdiary is an on-line calendar for which we have not written a complete interface, but have focused on the three most important functions: appointment, moving, and canceling appointments. Other functions, such as "to do" lists, window management, listing somebody's appointments, etc., can be dealt with in a similar fashion, and we plan to extend the interface to deal with them. At this point the application knowledge base is very simple. It consists of rules that say how to interpret the data given by the semantic interpreter, for instance the rules for formatting parameters and renaming slots (e.g. event_duration → duration). Such rules are necessary, if the distinction between application and domain knowledge is to be maintained.

The domain knowledge base: This has two kinds of facts: (1) background ontology, i.e., is, basic facts about time and places, and (2) linguistic knowledge associated with the domain. The former includes such obvious facts as the number of days in a month, which month follows the other, that offices are places etc. The latter includes facts about how the language is used. For example, the *filters* saying that places do not modify people, so that *I want to meet my manager in the cafeteria* can be unambiguously parsed, with "cafeteria" being a meeting place, and not an attribute of the manager.

The organization of knowledge: The issue of the organization of knowledge has been discussed at length in [8] and [9] and the formal model developed there is applicable in the present context. At this point, however, this formal model has only been implemented very crudely. Still the model is worth briefly discussing, because the conceptual distinctions made guide our work and have important practical consequences. The most important thing about it is that we discard the model of background knowledge as a logical theory, and replace it by a model consisting of collection of theories and mechanisms for putting them together depending on circumstances. Thus, the usual, two-part logical structures, consisting of a *metalevel* and an *object level*, are augmented by a third level — a *referential level*. The referential level is a partially ordered collection of theories; it encodes background knowledge in a way resembling a dictionary or an encyclopedia.[2]

Parser, construction grammar and linguistic knowledge

Parser: The parser does not produce (syntactic) structural descriptions of sentences. Instead, it computes meaning representations. For example, it converts adjuncts directly into attributes of place, time, participant etc., once they can be computed, and thus the message of the sentence does not contain any information about how these attributes where expressed or about the attachment of PPs that appear in it. For example, the sentence *I want you to arrange a conference in my office at 5* is analyzed as *sent(assert, svoc)*, an assertive sentence consisting of a subject, a verb, an object and a complement.

[1] These are constructions we used in MINCAL. In addition in various experiments we have used a few dozen other constructions, e.g. those covering "open idioms" (see Section 4).

[2] As usual, current situations are described on the object level, and the metalevel is a place for rules that can eliminate some of the models permitted by the object level and the referential level.

Google Exhibit 1020
Page 0055

The latter and the message of the imperative that is passed to *sent(assert, svoc)* does not contain any structural information about the attachment of the PPs. This message is combined with the messages of the verb and the noun, yielding

```
[  den want(other_agent)]
 [  agent hearer]
 [  mental_agent
   [  [  type person]
      [  den speaker]
[  [  action
      [  [  den arrange]
   [  action_object
         [  type event]
         [  den conference]
         [  number 1]
         [  mods
            [  [  det a]
               [  pp_msg
                  [  [  prep at]
                     [  type time(hour)]
                     [  den
                        [  [  hour
                              [  5 am_or_pm]
                           [  minute 0]
                  [  [  prep in]
                     [  type place ]
                     [  den office]
                     [  mods
                        [  [  det my]
```

This result of parsing is then interpreted by the domain interpreter to produce:

```
***Slots:
[  [  action_name schedule]
   [  event_name
      [  a conference]
   [  event_time
      [  [  minute 0]
         [  hour
            [  5 am_or_pm]
   [  event_place
      [  my office]
```

Application-specific defaults then produce yet another interpretation where, in addition to filling the slots of Xdiary, [hour [5 am_or_pm]] is interpreted as [hour [17]].

The parser is a chart parser, working left to right, with no lookahead. The grammar is L-attributed, i.e., has has both synthesized and inherited attributes, but each inherited attribute depends only on inherited attributes of the parent or attributes of the sisters to the left. Hence, although the parser does not have a lookahead step at present, such a step can be added following [2].

4 Comparisons with related work

Linguistic arguments for constructions-based grammars has been worked out chiefly by Ch. Fillmore and his colleagues (cf. [3]). Their motivation for advocating such an approach comes from the fact that typical generative theories of grammar cannot deal properly with open idioms illustrated by constructions such as:

The more carefully you work, the easier it will get.

Why not fix it yourself?

Much as I like Ronnie, I don't approve of anything he does.

It's time you brushed your teeth.

Him be a doctor?

The same is true about even so-called robust parsers of English. The reason for this failure can be attributed to the fact that expressions like these "exhibit properties that are not fully predictable from independently known properties of its lexical makeup and its grammatical structure" – [3], p.511. However we do not need a list of "strange" constructions to conclude that thoroughly integrating syntax with semantics and pragmatics could provide us with a better handle on natural language understanding. On a closer examination "normal" constructions exhibit enough complexity to warrant the new approach (see [10] for details).

Jurafsky [4] has independently come up with a proposal for a computable grammar of constructions. We compare our work with his in [10]. Here, we limit ourselves to a few remarks. What is common in both approaches is the centrality of the concept of grammatical construction as a data structure that represents lexical, semantic and syntactic knowledge. However, there are important differences between the two formalisms. First, the actual data structures used to represent constructions are different. The most important difference has to do with the presence of the *context* field in our version of the construction grammar. This allows us to account for the importance of pragmatics in representing many constructions, and to deal with discourse constructions.

Secondly, while Jurafsky acknowledges the need for abstract constructions (pp.43-51), his abstract constructions (*weak constructions*) are not first class citizens — they are defined only extensionally, by specifying the set of constructions they abstract over, and their abstract meaning (e.g. **entity** for NOUN). They are used to simplify descriptions of constituents of other constructions. However, because they do not have a separate *vehicle* part, they cannot be used to assign default meanings. For instance, since *verb* is defined as a collection of all verbs *is + read + cancel + know + look-up + ...*, it cannot be assigned a feature *action_verb* without introducing a contra-

Google Exhibit 1020
Page 0056

diction – its semantics is therefore given as RELATION/PROCESS. For us the important feature of "abstract" constructions is not that they simplify descriptions of other constructions, but that they have default meanings. (A similar critique of [5] can be found in [10]).

5 Summary of results

Our approach to NLU is based both on linguistic arguments and on our dissatisfaction with the state of the art. State of the art systems typically are too "syntax-driven", failing to take context into account in determining the intended meaning of sentences. A related further weakness is that such systems are typically "sentence oriented", rather than "conversation/discourse oriented". In our view, this makes even the most robust systems "brittle" and ultimately impractical.

To test whether a construction-based approach is feasible built a "complete" working system that would include a representation for constructions. To do this, we focused on the "calendar domain", a domain with limited complexity and simple but not uninteresting semantics. We have chosen to deal with simple actions, and not e.g. with question answering, where deeper understanding would be necessary. [3]

Our contributions:

1. We have proposed *a new kind of grammar* – computable construction grammars, which are neither semantic, nor syntactic. Instead, their "productions" combine lexical, syntactic, semantic and pragmatic information. [4]
2. We have described *data structures for constructions*, and have shown that they can be effectively used by the parser. Note that the same data structure is used to encode the lexicon and the "syntactic" forms.
3. We have shown *how to parse with constructions.* We have implemented a simple chart parsing algorithm, which can be easily extended to an Eearly-like parser, as long as the construction grammar remains L-attributed. We have found that even a simple parser of construction can be quite efficient. This is partly due to the fact that it does not require copying of all syntactic and semantic information from daughters to mothers; the goal of parsing consists in producing an interpretation, and structural information can be discarded once an interpretation of a phrase is produced. It is also worth emphasizing that invoking domain semantics drastically reduces the number of parses constructed.
4. We have proposed *a modular architecture for NL interfaces* based on the division between linguistic knowledge, domain knowledge base, and application knowledge base. Based on our experience, we believe that this architecture should work in general for speech-enabled interfaces for restricted domains.

[3] We have also thought about another possibility, that is, enhancing an IR system, e.g. with the understanding of date expressions.

[4] In what sense are they "computable"? Although this adjective might suggest a formal model with computational complexity results, etc., what we have in mind is pretty trivial: (1) the system actually computes the messages of grammatical construction; (2) the grammars and constructions are well defined data structures, and parsing (combining all associated constructions in all possible ways) is decidable.

References

[1] *Computational Linguistics.* Special Issue on the Lexicon, 1987. 13: 3-4.

[2] Nelson Correa. An extension of earley's algorithm for s- and l- attributed grammars. In *Proc. Intl. Conf On Current Issues in Computational Linguistics*, pages 248–261. Penang, Malaysia, 1991.

[3] Charles J. Fillmore, Paul Kay, and Mary Catherine O'Connor. Regularity and idiomaticity in grammatical constructions. *Language*, 64(3):501–538, 1988.

[4] D. Jurafsky. *An On-line Computational Model of Sentence Interpretation.* PhD thesis, University of California, Berkeley, 1992. Report No. UCB/CSD 92/676.

[5] Paul Kay. Even. *Linguistics and Philosophy*, 13(1):59–111, 1990.

[6] M. McCord, Arendse Bernth, Shalom Lappin, and Wlodek Zadrozny. Natural language processing technology based on slot grammar. *International Journal on Artificial Intelligence Tools*, 1(2):229–297, 1992.

[7] I. Melcuk. *Dependency Syntax: Theory and Practice.* State University of New York Press, Albany, NY, 1988.

[8] Wlodek Zadrozny. Reasoning with background knowledge – a three-level theory. *Computational Intelligence*, 10(2), 1994. To appear.

[9] Wlodek Zadrozny and Karen Jensen. Semantics of paragraphs. *Computational Linguistics*, 17(2):171–210, 1991.

[10] Wlodek Zadrozny and Alexis Manaster-Ramer. The significance of constructions. *Submitted to Computational Linguistics*, 1994.

Google Exhibit 1020
Page 0057

©COLING 94 Organizing Committee
All rights reserved. No part of this publication may be reproduced, stored in retrieval system or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior notice.

無断転載を禁ず。

6009 /007

Google Exhibit 1020
Page 0058

Appendix 1006B

COLING 94

The 15th International Conference on Computational Linguistics

PROCEEDINGS

Vol. II

August 5 – 9, 1994

Kyoto, Japan

GE

Google Exhibit 1020
Page 0060

COLING 94

The 15th International Conference on Computational Linguistics

PROCEEDINGS

Vol. II

August 5 – 9, 1994

Kyoto, Japan

Google Exhibit 1020
Page 0061

Table of Contents (Vol. I)

Machine Translation

Improvement in Customizability Using Translation Templates ---------- 25
S. Kinoshita, A. Kumano, H. Hirakawa (Toshiba Corp.)
Countability and Number in Japanese to English Machine Translation ---------- 32
F. Bond, K. Ogura, S. Ikehara (NTT Communication Science Labs.)
A Method for Distinguishing Exceptional and GeneraL Examples in Example-based Transfer Systems ---------- 39
H. Watanabe (IBM Japan, Ltd.)
Interpreting Compounds for Machine Translation ---------- 45
B. Gawronska, A. Nordner, C. Johansson, C. Willners (University of Lund)
Towards Machine Translation Using Contextual Information ---------- 51
T. Cornish, K. Fujita, R. Sugimura (Matsushita Electric Industrial Co., Ltd.)
Two Methods for Learning ALT-J/E Translation Rules from Examples and a Semantic Hierarchy ---------- 57
H. Almuallim, Y. Akiba, T. Yamazaki, A. Yokoo, S. Kaneda
(King Fahd Univ. of Petroleum and Minerals/NTT Communication Sci. Labs)
A Bidirectional, Transfer-driven Machine Translation System for Spoken Dialogues ---------- 64
Y. Sobashima, O. Furuse, S. Akamine, J. Kawai, H. Iida
(ATR Interpreting Telecommunication Research Laboratories)
Treating 'Free Word Order' in Machine Translation ---------- 69
R. Steinberger (UMIST)
Building an MT Dictionary from Parallel Texts Based on Linguistic and Statistical Information ---------- 76
A. Kumano, H.Hirakawa (Toshiba Corp.)
Dilemma - An Instant Lexicographer ---------- 82
H. Karlgren, J. Karlgren, M. Nordström, P. Pettersson, B. Wahrolén
(Swedish Institute of Computer Science)
Portable Knowledge Sources for Machine Translation ---------- 85
K. Takeda (IBM Research)
Coping with Ambiguity in a Large-Scale Machine Translation System ---------- 90
K. L. Baker, A. M.Franz, P. W.Jordan, T. Mitamura, E. H. Nyberg,3rd
(Carnegie Mellon University)
Evaluation Metrics for Knowledge-Based Machine Translation ---------- 95
E. H. Nyberg, 3rd, T. Mitamura, J. G.Carbonell (Carnegie Mellon University)
A Matching Technique in Example-Based Machine Translantion ---------- 100
L. Cranias, H. Papageorgiou, S. Piperidis (Institute for Language & Speech Processing, Greece)
Constituent Boundary Parsing for Example-Based Machine Translation ---------- 105
O. Furuse, H. Iida (ATR Interpreting Telecommunications Research Labs.)
The JaRAP Experimental System of Japanese-Russian Automatic Translation ---------- 112
L. S. Modina, Z. M. Shalyapina (Russian Academy of Sciences)
Perspectives of DBMT For Monolingual Authors on the Basis of LIDIA-1, an Implemented Mock-up ---------- 115
H. Blanchon (GETA–IMAG)

Google Exhibit 1020
Page 0062

Modals as a Problem for MT ---------- 120
B. Sigurd, B. Gawrónska (Lund University)
Two Types of Adaptive MT Environments ---------- 125
S. Nirenburg, R. Frederking, D. Farwell, Y. Wilks
(Carnegie Mellon Univ./New Mexico State Univ./Univ. of Sheffield)
An English-to-Korean Machine Translator:MATES/EK ---------- 129
K.-S. Choi, S. Lee, H. Kim, D.-B. Kim, C. Kweon, G. Kim
(Korea Advanced Institute of Science and Technology)
Interlanguage Signs and Lexical Transfer Errors ---------- 134
Atle Ro (University of Bergen)

Morphology & Tagging

Morphology with a Null-Interface ---------- 141
H. Trost, J. Matiasek (Austrian Research Institute for Artificial Intelligence)
Automatic Model Refinement - with an Application to Tagging ---------- 148
Y.-C. Lin, T.-H. Chiang, K.-Y. Su (National Tsing Hua University)
Disambiguation of Super Parts of Speech (or SUPERTAGS):Almost Parsing ---------- 154
A. K. Joshi, B.Srinivas (University of Pennsylvania)
Probabilistic Tagging with Feature Structures ---------- 161
A. Kempe (Univ. of Stuttgart)
A Part-of-Speech-Based Alignment Algorithm ---------- 166
K.-H. Chen (National Taiwan University)
Part-of-Speech Tagging with Neural Networks ---------- 172
H. Schmid (Univ. of Stuttgart)
The Rumors System of Russian Synthesis ---------- 177
M. I. Kanovich, Z. M.Shalyapina (Russian Academy of Sciences)
Multi-Tape Two-Level Morphology: A Case Study in Semitic Non-linear Morphology -------- 180
G. A. Kiraz (University of Cambridge)
An Evaluation to Detect and Correct Erroneous Characters Wrongly Substituted, Deleted and Inserted in Japanese and English Sentences Using Markov Models ---------- 187
T. Araki, S. Ikehara, N. Tsukahara, Y. Komatsu (Fukui Univ./NTT Communication Sci. Labs.)
An Efficienct Treatment of Japanese Verb Inflection for Morphological Analysis ---------- 194
T. Hisamitsu, Y. Nitta (ARL, Hitachi Ltd.)
*A Stochastic Japanese Morphological Analyzer Using a Forward-DP Backward-A*N-Best Search Algorithm* ---------- 201
M. Nagata (NTT Network Information Systems Labs.)
Backtracking-Free Dictionary Access Method for Japanese Morphological Analysis -------- 208
H. Maruyama (IBM Research)
Modularity in a Connectionist Model of Morphology Acquisition ---------- 214
M. Gasser (Indiana University)
Syllable-Based Model for the Korean Morphology ---------- 221
S.-S. Kang, Y. T. Kim (Hansung Univ./Seoul National Univ.)
Segmenting a Sentence into Morphemes Using Statistic Information between Words ------ 227
S. Nobesawa, J. Tsutsumi, T. Nitta, K. Ono, S. D. Jiang, M. Nakanishi (Keio University)
'Derivational' Paradigms in Morphonology ---------- 234
V. Pirrelli, S. Federici (ILC-CNR)

Lexicon

An Architecture for a Universal Lexicon. A Case Study on Shared Syntactic Information in Japanese, Hindi, Bengali, Greek, and English ---------- 243
N. Nomura, D. A. Jones, R. C. Berwick (Massachusetts Institute of Technology)
Adjuncts and the Processing of Lexical Rules ---------- 250
G. V. Noord, G. Bouma (BCN RUG)
Noun Phrasal Entries in the EDR English Word Dictionary ---------- 257
A. Koizumi, M.Arioka, C.Harada, M.Sugimoto
(Japan Electric Dictionary Research Institute)
KASSYS: A Definition Acquisition System in Natural Language ---------- 263
P. Hernert (L.I.F.O., Research Lab. de Vinci)
Comlex Syntax:Building a Computational Lexicon ---------- 268
R. Grishman, C. Macleod, A. Meyers (New York University)
Word Knowledge Acquisition, Lexicon Construction and Dictionary Compilation ---------- 273
A. Sanfilippo (SHARP Laboratories of Europe, Ltd.)
Interlingual Lexical Organisation for Multilingual Lexical Databases in NADIA ---------- 278
G. Sérasset (GETA IMAC-campus)
Manipulating Human-Oriented Dictionaries with Very Simple Tools ---------- 283
J. Gaschler, M. Lafourcade (GETA, IMAG-campus)
Towards Linguistic Knowledge Discovery Assistants:Application to Learning Lexical Properties of Chinese Characters ---------- 287
G. Fafiotte, F. Tcheou (GETA, IMAG)
Logic Compression of Dictionaries for Multilingual Spelling Checkers ---------- 292
B. Meddeb-Hamrouni (GETA, IMAG-campus)
Construction of a Bilingual Dictionary Intermediated by a Third Language ---------- 297
K. Tanaka, K. Umemura (Univ. of Tokyo/NTT Basic Research Labs.)
Co-occurrence Vectors from Corpora vs. Distance Vectors from Dictionaries ---------- 304
Y. Niwa, Y. Nitta (ARL, Hitachi Ltd.)
Analysis of Scene Identification Ability of Associative Memory with Pictorial Dictionary ---------- 310
T. Tsunoda, H. Tanaka (University of Tokyo)

Generation

Anticipating the Reader's Problems and the Automatic Generation of Paraphrases ---------- 319
N. Lenke (Univ. of Duisburg)
TGE:Tlinks Generation Environment ---------- 324
A. Ageno, F. Ribas, G. Rigau, H. Rodríguez, A. Samiotou (Univ. Politècnica de Catalunya)
Planning Argumentative Texts ---------- 329 ✓
X. Huang (Universität des Saarlandes)
A Hybrid Approach to the Automatic Planning of Textual Structures ---------- 334
G. Zhu, N. Shadbolt (Nottingham University)
Generating Multilingual Documents from a Knowledge Base:The TECHDOC Project ---------- 339
D. Rosner, M. Stede (FAW Ulm)
Abstract Generation Based on Rhetorical Structure Extraction ---------- 344
K. Ono, K. Sumita, S. Miike (Toshiba Corp.)
The Correct Place of Lexical Semantics in Interlingual MT ---------- 349
L. Levin, S. Nirenburg (Carnegie Mellon University)

Google Exhibit 1020
Page 0064

Default Handling in Incremental Generation ---------- 356
K. Harbusch, G. Kikui, A. Kilger (German Research Center for Artificial Interlligence)
English Generation from Interlingua by Example-Based Method ---------- 363
E. Komatsu, J. Cui, H. Yasuhara (Japan Electric Dictionary Res. Inst.)
On Lexically Biased Discourse Organization in Text Generation ---------- 369
L. Wanner (University of Stuttgart)

Parsing

Concurrent Lexicalized Dependency Parsing:The Parse Talk Model ---------- 379
N. Bröker, U. Hahn, S. Schacht (Freiburg University)
Notes on LR Parser Design ---------- 386
C. Samuelsson (Swedish Institute of Computer Science)
Parsing a Flexible Word Order Language ---------- 391
V. Pericliev, A. Grigorov (Institute of Mathematics with Computing Center, Bulgaria)
Parsing as Tree Traversal ---------- 396
D. Gerdemann (Universität Tübingen)
LR(k)-Parsing of Coupled-Context-Free Grammars ---------- 401
G. Pitsch (Universität des Saarlandes)
Constructing Lexical Transducers ---------- 406
L. Karttunen (Rank Xerox Research Centre, Grenoble)
An HPSG Parser Based on Description Logics ---------- 412
J. Joachim Quantz (Technische Universität Berlin)
An Efficienct Parser Generator for Natural Language ---------- 417
M. Ishii, K. Ohta, H. Saito (Fujitsu Inc./Apple Technology, Inc./Keio Univ.)
Exploring the Role of Punctuation in Parsing Natural Text ---------- 421
B. E.M. Jones (University of Edinburgh)
The "Whiteboard" Architecture:A Way to Integrate Heterogeneous Components of NLP Systems ---------- 426
C. Boitet, M. Seligman (GETA, IMAG/ATR Interpreting Telecommunications Res. Labs)
Two Parsing Algorithms by Means of Finite State Transducers ---------- 431
E. Roche (Mitsubishi Electric Research Labs.)
DISCO–An HPSG-Based NLP System and its Application for Appointment Scheduling --- 436
H. Uszkoreit, R. Backofen, S. Busemann, A. K. Diagne, E. A.Hinkelman, W. Kasper, B. Kiefer, H.-U. Krieger, K. Netter, G. Neumann, S. Oepen, S. P.Spackman (DFKI)
A Corpus-Based Learning Technique for Building a Self-Extensible Parser ---------- 441
R.-L. Liu, V.-W. Soo (National Tsing-Hua University)
A "Not-so-shallow" Parser for Collocational Analysis ---------- 447
R. Basili, M.T.Pazienza, P.Velardi (Università di Roma)
A Modular Architecture for Constraint-based Parsing ---------- 454
F. Barthélemy, F. Rouaix (Universidade Nova de Lisboa/INRIA Rocquencourt)
Minimal Change and Bounded Incremental Parsing ---------- 461
M. Wirén (Universität des Saarlandes)
Emergent Parsing and Generation with Generalized Chart ---------- 468
K. Hashida (Electrotechnical Laboratory, Japan)
Syntactic-Head-Driven Generation ---------- 475
E. Konig (Institute for Computational Linguistics, Germany)

PRINCIPAR-An Efficient, Broad-coverage, Principle-based Parser ---------- 482
D. Lin (University of Manitoba)
Concurrent Lexicalized Dependency Parsing: A Behavioral View on Parse Talk Events --- 489
S. Schacht, U. Hahn, N. Broker (Freiburg University)
Parsing Turkish Using the Lexical Functional Grammar Formalism ---------- 494
Z. Güngördü, K. Oflazer (University of Edinburgh/Bilkent Univ.)
LHIP: Extended DCGs for Configurable Robust Parsing ---------- 501
A. Ballim, G. Russell (ISSCO, University of Geneva)
Categorial Grammar and Discourse Representation Theory ---------- 508
R. Muskens (Institute for Language Technology and Artificial Intelligence)
Towards Automatic Extraction of Monolingual and Bilingual Terminology ---------- 515
B. Daille, E. Gaussier, J.-M. Langé
(Talana Univ. Paris 7/Centre Scientifique IBM France)

Computational Linguistics

FAX: An Alternative to SGML ---------- 525
K. W. Church, W. A. Gale, J. I.Helfman, D. D.Lewis (AT&T Bell Labs.)
Referring to World Objects with Text and Pictures ---------- 530
E. André, T. Rist (DFKI)
A Two-level Morphological Analysis of Korean ---------- 535
D.-B. Kim, S.-J. Lee, K.-S. Choi, G.-C. Kim
(Korea Advanced Institute of Science and Technology)
Character-based Collocation for Mandarin Chinese ---------- 540
C.-R. Huang, K.-J. Chen, Y.-Y. Yang (Academia Sinica/National Taiwan Univ.)
Lexical Knowledge Representation in an Intelligent Dictionary Help System ---------- 544
E. Agirree, X.Arregi, X.Artola, A.Díaz de Iiarraza, K.Sarasola (Univ. of Basque Country)
Reversible Resolution with an Application to Paraphrasing ---------- 551
M. Hurst (The University of Sheffield)
Classifier Assignment by Corpus-Based Approach ---------- 556
V. Sornlertlamvanich, W. Pantachat, S. Meknavin (Linguistics and Knowledge Science Lab., Ministry of Science Technology and Environment, Thailand)

Corpus-based NLP

Annotating 200 Million Words:The Bank of English Project ---------- 565
T. Järvinen (University of Helsinki)
Restructuring Tagged Corpora with Morpheme Adjustment Rules ---------- 569
T. Tashiro, N. Uratani, T. Morimoto (ATR Interpreting Telecommunications Research Labs.)
Encoding Standards for Large Text Resources:The Text Encoding Initiative ---------- 574
N. Ide (Vassar College)
INTEX:A Corpus Processing System ---------- 579
M. D. Silberztein (Université Paris 7)
An IBM-PC Environment for Chinese Corpus Analysis ---------- 584
R. W. P. Luk (City Polytechnic of Hong Kong)
Multext:Multilingual Text Tools and Corpora ---------- 588
N. Ide, J. Véronis (CNRS & Universite de Provence)

Google Exhibit 1020
Page 0066

A Parser Coping with Self-Repaired Japanese Utterance and Large Corpus-Based Evaluation ---------- 593
Y. Sagawa, N. Ohnishi, M. Sugie (Nagoya University)
A Tool for Collecting Domain Dependent SortaL Constraints from Corpora ---------- 598
F. Andry, M. Gawron, J. Dowding, R. Moore (SRI International/CAP GEMINI Innovation)
Building a Lexical Domain Map from Text Corpora ---------- 604
T. Strzalkowski (New York University)
A New Method of N-gram Statistics for Large Number of n and Automatic Extraction of Words and Phrases from Large Text Data of Japanese ---------- 611
M. Nagao, S. Mori (Kyoto University)
Non-directionality and Self-Assessment in an Example-based System Using Genetic Algorithms ---------- 616
Y. Lepage (Universiti Sains Malaysia)
CLAWS4:The Tagging of the British National Corpus ---------- 622
G. Leech, R. Garside, M. Bryant (Lancaster University)
Syntactic Analysis of Natural Language Using Linguistic Rules and Corpus-Based Patterns ---- 629
P. Tapanainen, T. Järvinen (Rank Xerox Research Centre/Univ. of Helsinki)

Table of Contents (Vol. II)

Semantics

Word Sense Acquisition for Multilingual Text Interpretation ---------- 665
P. S. Jacobs (GE Research and Development Center)
A System of Verbal Semantic Attributes Focused on the Syntactic Correspondence between Japanese and English ---------- 672
H. Nakaiwa, A. Yokoo, S. Ikehara (NTT Communication Science Labs.)
Semantics of Complex Sentences in Japanese ---------- 679
H. Nakagawa, S. Nishizawa (Yokohama National University)
Content Characterization Using Word Shape Tokens ---------- 686
P. Sibun, D. S.Farrar (Fuji Xerox Palo Alto Lab.)
The Nature of Near-synonymic Relations ---------- 691
C. DiMarco (University of Waterloo)
Virtual Polysemy ---------- 696
A. Sanfilippo, K. Benkerimi, D. Dwehus (Sharp Labs. of Europe)
Building a Windows-based Bilingual Functional Semantic Processor ---------- 701
J. J. Webster (City Polytechnic of Hong Kong)
On the Proper Role of Coercion in Semantic Typing ---------- 706
J. Pustejovsky, P. Bouillon (Brandeis Univ./Univ. of Geneva)
Word Sense Ambiguation:Clustering Related Senses ---------- 712
W. B. Dolan (Microsoft Corp.)
A Best-match Algorithm for Broad-coverage Example-based Disambiguation ---------- 717
N. Uramoto (IBM Japan Ltd.)
Drawing Pictures with Natural Language and Direct Manipulation ---------- 722
M. Hiyoshi, H. Shimazu (NEC Corp.)
Verbal Case Frame Acquisition from a Bilingual Corpus:Gradual Knowledge Acquisition --- 727
H. Tanaka (NHK Science & Technical Research Labs.)
An Empirical Study on the Generation of Zero Anaphors in Chinese ---------- 732
C.-L. Yeh, C. Mellish (University of Edinburgh)
Representing Information Need with Semantic Relations ---------- 737
A. S. Chakravarthy (MIT Media Lab.)
Generalizing Automatically Generated Selectional Patterns ---------- 742
R. Grishman, J. Sterling (New York University)
Incremental Interpretation: Applications, Theory, and Relationship to Dynamics Semantics 748
D. Milward, R. Cooper (University of Edinburgh)
Word Sense Disambiguation and Text Segmentation Based on Lexical Cohesion ---------- 755
M. Okumura, T. Honda (Japan Advanced Institute of Science and Technology)
Automatic Recognition of Verbal Polysemy ---------- 762
F. Fukumoto, J. Tsujii (UMIST)
An Experiment on Learning Appropriate Selectional Restrictions from a Parsed Corpus --- 769
F. R. Framis (Universitat Politècnica de Catalunya)

Google Exhibit 1020
Page 0068

A Discrete Model of Degree Concept in Natural Language ---------- 775
S. Kamei, K. Muraki (NEC Corp.)
Algorithm for Automatic Interpretation of Noun Sequences ----------782
L. Vanderwende (Microsoft Corp.)
A Treatment of Functional Definite Descriptions ----------789
H. Wada (Intelligent Text Processing, Inc.)
Bottom-up Earley Deduction ----------796
G. Erbach (University of Saarland)
The Merged Upper Model:A Linguistic Ontology for German and English ----------803
R. Henschel, J. Bateman (GMD/IIPI)

Syntax

Modeling Dialogue by Functional Subcategorization ----------813
J.R. Zubizarreta Aizpuru, C. Jones (U.P.V.–E.H.U./Univ. of Aberdeen)
Dutch Cross Serial Dependencies in HPSG ----------818
G. Rentier (Tilburg University)
HPSG Lexicon without Lexical Rules ----------823
K. Oliva (Universiy of Saarland)
A Lexicon of Distributed Noun Representations Constructed by Taxonomic Traversal ------- 827
R. Sutcliffe, D. O'Sullivan, F. Meharg (University of Limerick)
Multi-modal Definite Clause Grammar ----------832
H. Shimazu, S. Arita, Y. Takashima (NEC Corp.)
Hypothesis Selection in Grammar Acquisition ----------837
M. Kiyono, J. Tsujii (UMIST)
Achieving Flexibility in Unification Formalisms ----------842
L. Strömbäck (Linköping University)
TWP:How to Assist English Production on Japanese Word Processor ----------847
K. Muraki, S. Akamine, K. Satoh, S. Ando (NEC Corp.)
Long-distance Dependencies and Applicative Universal Grammar ----------853
S. Shaumyan, F. Segond (Yale University/Rank Xerox Res. Centre)
A Reestimation Algorithm for Probabilistic Recursive Transition Network ----------859
Y. S. Han, K.-S. Choi (Korea Advanced Institute of Science and Technology)
Analysis of Japanese Compound Nouns Using Collocational Information ----------865
Y. Kobayashi, T. Tokunaga, H. Tanaka (Tokyo Institute of Technology)
Free-ordered CUG on Chemical Abstract Machine ----------870
S. Tojo (Mitsubishi Research Institute, Inc.)
Computing FIRST and FOLLOW Functions for Feature-theoretic Grammars ----------875
A. Trujillo (University of Cambridge)
Focus on "Only" and "Not" ----------881
A. Ramsay (University College Dublin)
Structure Sharing Problem and its Solution in Graph Unification ----------886
K. Kogure (NTT Basic Research Labs.)
TDL-A Type Description Language for Constraint-based Grammars ----------893
H.-U. Krieger, U. Schäfer (German Research Center for Artificial Intelligence)
On the Portability of Complex Constraint-based Grammars ----------900
C.J. Rupp, R. Johnson (IDSIA)

A Grammar Based Approach to a Grammar Checking of Free Word Order Languages ---- 906
V. Kubon, M. Plátek (Charles University)
Table-driven Neural Syntactic Analysis of Spoken Korea ---------- 911
W. Lee, G. Lee, J.-H. Lee (Phonag Institute of Science and Technology)
Universal Guides and Finiteness and Symmetry of Grammar Processing Algorithms ------ 916
M. Martinovic (New York University)
XTAG System-A Wide Coverage Grammar for English ---------- 922
C. Doran, D. Egedi, B. A. Hockey, B.Srinivas, M. Zaidel (University of Pennsylvania)
Weakly Restricted Stochastic Grammars ---------- 929
R. O. D. Akker, H. ter Doest (University of Twente)
Non-constituent Coordination:Theory and Practice ---------- 935
D. Milward (University of Edinburgh)
Hypothesis Scoring over Theta Grids Information in Parsing Chinese Sentences with Serial Verb Constructions ---------- 942
K. H.C. Lin, V.-W. Soo (National Tsing-Hua University)
An Efficient Syntactic Tagging Tool for Corpora ---------- 949
M. Zhou, C. Huang (Tsinghua University)
The Correct and Efficient Implementation of Appropriateness Specifications for Typed Feature Structures ----------956
D. Gerdemann, P. J. King (Universität Tübingen)
A Classification Method for Japanese Signs Using Manual Motion Descriptions ----------961
H. Adachi, K. Kamata (Utsunomiya University)

Speech & Phonology

Machine-readable Dictionaries in Text-to-speech Systems ----------971
J. L. Klavans, E. Tzoukermann (Columbia University)
Issues in Text-to-speech for French ----------976
E. Tzoukermann (AT&T Bell Labs.)
CHATR: A Generic Speech Synthesis System ----------983
A. W. Black, P. Taylor (ATR Interpreting Telecommunications Labs.)
Pause as a Phrase Demarcator for Speech and Language Processing ----------987
J. Hosaka, M. Seligman, H. Singer (ATR Interpreting Telephony Research Labs.)
The Parsody System: Automatic Prediction of Prosodic Boundaries for Text-to-speech --- 992
S. Minnis (BT Labs.)
Anytime Algorithms for Speech Parsing? ----------997
G. Görz, M. Kesseler (University of Erlangen-Nürnberg)
Towards a Proper Linguistic and Computational Treatment of Scrambling: An Analysis of Japanese ----------1002
S. Fong (NEC Research Institute, Inc.)
Phonological Derivation in Optimality Theory ----------1007
M. T. Ellison (University of Edinburgh)
A Grammar and a Parser for Spontaneous Speech ----------1014
M. Nakano, A. Shimazu, K. Kogure (NTT Basic Research Labs.)
Catching the Cheshire Cat ----------1021
C. Johansson (Lund University)

Google Exhibit 1020
Page 0070

Information Retrieval & Extraction

A Dutch to SQL Database Interface Using Generalized Quantifier Theory ---------- 1029
D. Speelman, G. Adriaens (University of Leuven)
A Methodology for Automatic Term Recognition ---------- 1034
S. Ananiadou (Manchester Metropolitan University)
Knowledge Extraction from Texts: A Method for Extracting Predicate-argument Structures from Texts ---------- 1039
F. Pugeault, P. Saint-Dizier, M.-G. Monteil (Université Paul Sabatier/EDF, D.E.R.)
Thesaurus-based Efficient Example Retrieval by Generating Retrieval Queries from Similarities ---------- 1044
T. Utsuro, K. Uchimoto, M. Matsumoto, M. Nagao
(Nara Institute of Science and Technology/Kyoto University)
Tools for Extracting and Structuring Knowledge from Texts ---------- 1049
A. Ogonowski, M. L. Herviou, E. Dauphin (GSI-ERLI/AEROSPATIALE CCR/EDF)
N-Gram Cluster Identification During Empirical Knowledge Representation Generation -- 1054
R. Collier (University of Sheffield)
Document Classification by Machine: Theory and Practice ---------- 1059
L. Guthrie, E. Walker (New Mexico State University)
Pattern Matching in the Textract Information Extraction System ---------- 1064
T. Kitani, Y. Eriguchi, M. Hara (Carnegie Mellon University)
Recognizinig Text Genres with Simple Metrics Using Discriminant Analysis ---------- 1071
J. Karlgren, D. Cutting (Swedish Institute of Computer Science/Apple Computer)
Bilingual Text Matching Using Bilingual Dictionary and Statistics ---------- 1076
T. Utsuro, H. Ikeda, M. Yamane, Y. Matsumoto, M. Nagao
(Nara Institute of Science and Technology/Kyoto University)
Towards a More User-friendly Correction ---------- 1083
D. Genthial, J. Courtin, J. Ménézo (LGI-IMAG)
Reverse Queries in DATR ---------- 1089
H. Langer (University of Osnabrück)
K-vec: A New Approach for Aligning Parallel Texts ---------- 1096
P. Fung, K. W. Church (Columbia University/AT&T Bell Labs.)

Discourse & Pragmatics

A Formal Representation of the Thematic-rhematic Structure of Sentences Based of a Typed A-calculus ---------- 1105
Y. Uetake (Tokyo University of Mercantile Marine)
Customizing and Evaluating a Multilingual Discourse Module ---------- 1109
C. Aone (Systems Research and Applications Corp. [SRA])
Consequence Relations in DRT ---------- 1114
S. Akama, Y. Nakayama (Teikyo Univ. of Technology/Nihon Unisys Ltd.)
Collaboration on Reference to Objects that are not Mutually Known ---------- 1118
P. G. Edmonds (University of Toronto)
Automatic Detection of Discourse Structure by Checking Surface Information in Sentences 1123
S. Kurohashi, M. Nagao (Kyoto University)

Extending DRT with a Focusing Mechanism for Pronominal Anaphora and Ellipsis Resolution ---------- 1128
J. Abraços, J. G. Lopes (CRIA/UNINOVA Faculdade de Ciências e Tecnologia)
Reference Resolution Using Semantic Patterns in Japanese Newspaper Articles ---------- 1133
T. Wakao (University of Sheffield)
Exploiting Reference Interaction in Resolving Temporal Reference ---------- 1138
K. Dohsaka (NTT Basic Research Labs.)
A Grammatico-statistical Approach to Discourse Partitioning ---------- 1145
T. Nomoto, Y. Nitta (Hitachi Ltd., ARL)
Centering in Japanese: A Step Towards Better Interpretation of Pronouns and Zero-pronouns ---------- 1151
S. Takada, N. Doi (Keio University)
Robust Method of Pronoun Resolution Using Full-text Information ---------- 1157
T. Nasukawa (IBM Research)
Towards a Dynamic Theory of Belief-sharing in Cooperative Dialogues ---------- 1164
H. Komatsu, N. Ogata, A. Ishikawa (Toin Corp./Univ. of Tsukuba/Sophia Univ.)
An Integrated Model for Anaphora Resolution ---------- 1170
R. Mitkov (Institute of Mathematics)
Breaking Down Rhetorical Relations for the Purpose of Analysing Discourse Structures -- 1177
J. Fukumoto, J. Tsujii (UMIST)
Presupposition & VP-Ellipsis ---------- 1184
J. Bos (Universität des Saarlandes)
Communicating with Multiple Agents ---------- 1191
E. A. Hinkelman, S. P. Spackman (DFKI)
A Rule-based Approach to Prepositional Phrase Attachment Disambiguation ---------- 1198
E. Brill, P. Resnik (Massachusetts Institute of Technology/Sun Microsystems Labs.)
Discourse and Deliberation: Testing a Collaborative Strategy ---------- 1205
M. A. Walker (Mitsubishi Electric Research Labs.)
A Bayesian Approach for User Modeling in Dialogue Systems ---------- 1212
T. Akiba, H. Tanaka (Tokyo Institute of Technology)

Google Exhibit 1020
Page 0072

Reserve Paper

World Class Discovery for Postprocessing Chinese Handwriting Recognition ---------- 1221
C.-H. Chang (Industrial Technology Research Institute)

A Simple Transformation for Offline-Parsable Grammars and Its Termination Properties -- 1226
M. Dymetman (Rank Xerox Research Center)

The Evaluation of Machine-Tractable Dictionaries ---------- 1231
C.-M. Guo, C.-N. Huang, J.-P. Gong, J. Li (Tsinghua University)

Discontinuity and the Lambek Calculus ---------- 1235
M. Hepple (University of Sheffield)

Lexical Functions and Machine Translation ---------- 1240
D. Heylen, K.G. Maxwell, M. Verhagen (OTS)

Chinese Segmentation Disambiguation ---------- 1245
W. Jin (New Mexico State University)

Typed Feature Structures as Descriptions ---------- 1250
P.J. King (Eberhard-Karls-Universität)

Portuguese Analysis with Tree Adjoining Grammars ---------- 1255
K.C. Kipper, V.L.S. de Lima (PUCRS)

Incremental Construction of a Lexical Transducer for Korean ---------- 1262
H.-C. Kwon, L. Karttunen (Pusan National University)

Dynamic Logic with Possible World ---------- 1267
R. Lu (Shanghai Jiao Tong University)

Humor-based Applications ---------- 1270
G. Prószéky, M. Pál, L. Tihanyi (Morphologic/OPKM Comp. Centre/Inst. for Linguistics)

Blending Segmentation with Tagging in Chinese Language Corpus Processing ---------- 1274
Z. Qiang, Y. Shiwen (Peking University)

Syllable-based Phonetic Transcription by Maximum Likelihood Methods ---------- 1279
R.A. Sharman (IBM UK Labs., Ltd.)

A Knowledge Acquisition and Management System for Morphological Dictionaries --------- 1284
P. Ten Hacken, S. Bopp, M. Domening, D. Holz, A. Hsiung, S. Pedrazzini
(Universitat Basel/Vrije Univ. Amsterdam/IDSIA)

NL Understanding with a Grammar of Constructions ---------- 1289
W. Zadrozny, M. Szummer, S. Jarecki, D.E. Johnson, L. Morgenstern (IBM)

NL Understanding with a Grammar of Constructions

Wlodek Zadrozny and Marcin Szummer and Stanislaw Jarecki
and David E. Johnson and Leora Morgenstern *
IBM Research Division
T.J.Watson Research Lab
Yorktown Heights, NY 10598 USA

Abstract

We present an approach to natural language understanding based on a *computable grammar of constructions*. A *construction* consists of a set of features of form and a description of meaning in a context. A grammar is a set of constructions. This kind of grammar is the key element of MINCAL, an implemented natural language speech-enabled interface to an on-line calendar system. The architecture has two key aspects: (a) the use of constructions, integrating descriptions of form, meaning and context into one whole; and (b) the separation of *domain knowledge* (about calendars) from *application knowledge* (about the particular on-line calendar).

1 Introduction: an overview of the system

We present an approach to natural language understanding based on a *computable grammar of constructions*. A *construction* consists of a set of features of form and a description of meaning in a context. A grammar is a set of constructions. This kind of grammar is the key element of MINCAL, an implemented natural language speech-enabled interface to an on-line calendar system.

The system consists of a NL grammar, a parser, an on-line calendar, a domain knowledge base (about dates, times and meetings), an application knowledge base (about the calendar), a speech recognizer, a speech generator.

In this paper we describe two key aspects of the system architecture: (a) the use of constructions, where instead of separating NL processing into the phases of syntax, semantics and pragmatics, we integrate descriptions of form, meaning and context into one whole, and use a parser that takes into account all this information (see [10] for details); (b) the separation of the *domain knowledge* (about calendars) and the *application knowledge* (about the particular on-line calendar).

*M. Szummer and S. Jarecki are also from MIT.

The dialogs

The system allows users to engage in dialogs like:

— *Schedule a meeting with Bob!*
— *At what time and date?*
— *On August 30th.*
— *At what time?*
— *At 8.*
— *Morning or afternoon?*
— *In the evening.*

The parser recognizes *Schedule a meeting with Bob* as an instance of *sent(imp)*, the imperative construction consisting of a verb and an NP, here *np(event)*. The context is used to prevent another reading in which *with Bob* modifies *schedule*, as in *Dance a tango with Bob!*. That is, a contextual rule is used which says that for calendar applications, people do not modify actions or places. Context also plays an important role in understanding answers, e.g. *At 8*. This is understood as a time expression (and not place or rate or something else) only because of the context.

The parameters of a meeting can be given in many ways, e.g. synonyms or different constructions can be used, users can include as many parameters in a sentence as they wish, and the parameters can be given in any order. As a result there are about 10,000 ways of scheduling meetings (with a given set of parameters).

How are the dialogs understood

With respect to parsing, grammars of constructions can be parsed like "standard" grammars, except that the set of features is richer. Given a string (representing a sentence, a fragment of a discourse or a paragraph), the parser assigns it a construction. From this viewpoint, the situation is similar to "regular" parsing, and the possible algorithms are similar. We have implemented a prototype chart parser for construction grammars, discussed further in Section 3.

But, clearly, having understood the sentence as a linguistic entity in isolation is not the ultimate goal. Here the message of an utterance must be understood in the context of an intended action. This is done in two steps. First, the system determines the intended

action and its parameters, using domain knowledge (meetings+time+places). Second, once all the parameters have been extracted from the dialog, the system executes the action. To do this, the program uses application-specific knowledge to translate the action and its parameters into a form that can be executed by the application (Xdiary).

2 Constructions as data structures

A construction is given by the matrix:

$$\begin{bmatrix} \mathbf{N} & : \textit{name_of_construction} \\ \begin{bmatrix} \mathbf{C} & : & \textit{context} \\ \mathbf{V} & : & \textit{structure} \\ \mathbf{M} & : & \textit{message} \end{bmatrix} \end{bmatrix}$$

The *vehicle* **V** consists of formulas describing presence (or perhaps absence) of certain *taxemes*, or features of form, within the structure of the construction. Such a structure is given by a list of subconstructions and the way they have been put together (in all our examples this is concatenation, but there are other possibilities, e.g. wrapping). The *context*, **C** , consists of a set of semantic and pragmatic constraints limiting the application of the construction. It can be viewed as a set of *preconditions* that must be satisfied in order for a construction to be used in parsing. The *message*, **M** , describes the meaning of the construction, via a set of syntactic, semantic and pragmatic constraints.

To make this concrete, let us consider a few examples. We begin with a simple "command construction" consisting of an action verb followed by its argument.

$$\begin{bmatrix} \mathbf{N} & : \textit{sent}(\textit{cmnd}, \textit{v.np}) \\ \begin{bmatrix} \mathbf{C} & : [< \textit{hr attends} >= \textit{sr}] \\ \mathbf{V} & : \begin{bmatrix} \textit{struc} = (V.NP) \\ < V\ \textit{cons_n} >= \textit{verb} \\ < V\ \mathbf{M}\ \ \textit{v_type} >= \textit{action_verb} \\ < NP\ \textit{cons_n} >= \textit{np} \end{bmatrix} \\ \mathbf{M} & : \begin{bmatrix} \textit{sem_cat} = \textit{command} \\ \textit{a_type} =< V\ \mathbf{M}\ \ \textit{sem_type} > \\ \textit{a_obj} =< NP\ \mathbf{M}\ \ \textit{sem_type} > \\ \textit{agent} = \textit{hr} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

The context of the construction describes all situations in which the the hearer *hr* (human or machine) is paying attention to the speaker *sr* (a "ready" state). The feature *struc* is a list of variables and/or words/tokens; it is used to describe the structure of a construction, and its role is similar to a rule in a generative grammar. (We will write names of variables in capital letters, e.g. *NP*, inside matrices of constructions). The attribute *cons_n* gives the name of a construction that could be assigned to a string. We use it here to say that the form of the construction can be described as a concatenation of two strings, of which one is a verb (construction) and the other an np (construction). Furthermore, the verb type $< V\ \mathbf{M}\ \ v_type >$ is "action_verb". (The expression $< V\ \mathbf{M}\ \ v_type >$ should be read "the *v_type* of the message of *V*").

The message **M** describes the meaning of the construction as that of a command in which the type of action is described by the meaning of the verb, and the object of the action is given by the meaning of the noun phrase. The attribute *sem_type* stands for the "semantic type" and we identify it currently with the word sense. Thus "erase the file" is understood as a command to *delete* the file, if $< \textit{erase}\ \mathbf{M}\ \ \textit{sem_type} >= \textit{delete}$, but "erase the picture" might refer to the type of action associated with *rub_out*. In both cases the hearer *hr* is supposed to be the agent of the action.

Constructions: from words to discourse

Words, phrases, and fragments of discourse can be analyzed as constructions. We view languages as collections of constructions which range from words to discourse. We claim that the same representation scheme can be used for all constructions.

The examples we are going to present have been developed with a specific purpose in mind, namely for scheduling calendar events. In other papers ([10] and [6]), we have presented examples showing that we can give a good descriptions of non-standard constructions. However, in either case descriptions of meanings and contexts are general, and hence applicable to other tasks.

We now turn our attention to words. The verb "cancel" can be represented as follows:

$$\begin{bmatrix} \mathbf{N} & : \textit{verb}(\textit{cancel}) \\ \begin{bmatrix} \mathbf{C} & : & \begin{bmatrix} \textit{lang_code} & = & \textit{english} \\ \textit{lang_channel} & = & \textit{text} \end{bmatrix} \\ \mathbf{V} & : & \textit{struc} = (\textit{cancel}) \\ \mathbf{M} & : & \begin{bmatrix} \textit{cat} & = & \textit{verb} \\ \textit{sem_type} & = & \textit{delete} \\ \textit{v_type} & = & \textit{action_verb} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Notice that even simple words require context to be (properly) interpreted. In **C** we say that English text is expected (but in other cases it could also be French text, or French speech, etc.). Some aspects of context do not have to be explicitly specified and can be replaced by defaults.

Although the vehicle and the message are both very simple in this example, the simplicity of the message is a result of deliberate simplification. We have restricted it to the specification of the semantic type, identified with one sense of the word, and to describing the verb type of "cancel" as a verb of action. Notice that the other sense of "cancel" – "offset, balance out" – would appear in another entry.

Of course, in reality, the lexical meaning of any word is a much more complicated matter [1]. For instance, in our lexicon the messages of words may contain many of the attributes that appear in the

Google Exhibit 1020
Page 0075

explanatory combinatorial dictionary of Melcuk [7].

Discourse constructions: To illustrate discourse constructions, we consider the following dialog:

Have you arranged the room yet?
No, but I'll do it right away.

We view the pattern of the answer *no.but.S* as a discourse construction. It can represented by the following array of features:

$$\begin{bmatrix} \mathbf{N} & : sent(assrt, no.but.S) \\ \begin{bmatrix} \mathbf{C} & : [< p_utter\ cons_n >= sent(ques, *)] \\ \mathbf{V} & : \begin{bmatrix} struc = (no.but.S) \\ < S\ cons_n >= sent(assrt, *) \end{bmatrix} \\ \mathbf{M} & : \begin{bmatrix} < p_sent\ truth_value >= 0 \\ < S\ \mathbf{M} > \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

As we can see, the construction applies only in the context of a previously asked question, and its message says that the answer to the question is negative, after which it elaborates the answer with a sentence S.

3 System Architecture

The parts

MINCAL consists of a NL grammar, a parser, a domain knowledge base (about dates, times and meetings), an on-line calendar (Xdiary), an application knowledge base (about Xdiary), a continuous speech recognizer (IBM, ICSS), a speech generator (Speech Plus, Text to Speech Converter), and the interfaces.

At present, the grammar consists of a few hundred lexical constructions, and about 120 "productions", i.e. constructions describing combinations of other constructions. [1] It covers the basic forms of assertive sentences, but it emphasizes commands. Thus a command can, for example, be given either by v.np (also with "please", or "kindly"), or by an assertive sentence recognized as an indirect speech act ("I'd like to ...", "Leora wants you to ...", etc.). The next large group of constructions covers PPs, with particular emphasis on time and places. Finally, it covers a few discourse constructions, since it is important to deal with sentence fragments in dialogs, e.g. understanding "evening" as "in the evening", when it is an answer to the question "when?".

The interaction of the modules

The calendar and the application knowledge base: Xdiary is an on-line calendar for which we have not written a complete interface, but have focused on the three most important functions: appointment, moving, and canceling appointments. Other functions, such as "to do" lists, window management, listing somebody's appointments, etc., can be dealt with in a similar fashion, and we plan to extend the interface to deal with them. At this point the application knowledge base is very simple. It consists of rules that say how to interpret the data given by the semantic interpreter, for instance the rules for formatting parameters and renaming slots (e.g. event_duration → duration). Such rules are necessary, if the distinction between application and domain knowledge is to be maintained.

The domain knowledge base: This has two kinds of facts: (1) background ontology, i.e., is, basic facts about time and places, and (2) linguistic knowledge associated with the domain. The former includes such obvious facts as the number of days in a month, which month follows the other, that offices are places etc. The latter includes facts about how the language is used. For example, the *filters* saying that places do not modify people, so that *I want to meet my manager in the cafeteria* can be unambiguously parsed, with "cafeteria" being a meeting place, and not an attribute of the manager.

The organization of knowledge: The issue of the organization of knowledge has been discussed at length in [8] and [9] and the formal model developed there is applicable in the present context. At this point, however, this formal model has only been implemented very crudely. Still the model is worth briefly discussing, because the conceptual distinctions made guide our work and have important practical consequences. The most important thing about it is that we discard the model of background knowledge as a logical theory, and replace it by a model consisting of collection of theories and mechanisms for putting them together depending on circumstances. Thus, the usual, two-part logical structures, consisting of a *metalevel* and an *object level* , are augmented by a third level -- a *referential level.* The referential level is a partially ordered collection of theories; it encodes background knowledge in a way resembling a dictionary or an encyclopedia.[2]

Parser, construction grammar and linguistic knowledge

Parser: The parser does not produce (syntactic) structural descriptions of sentences. Instead, it computes meaning representations. For example, it converts adjuncts directly into attributes of place, time, participant etc., once they can be computed, and thus the message of the sentence does not contain any information about how these attributes where expressed or about the attachment of PPs that appear in it. For example, the sentence *I want you to arrange a conference in my office at 5* is analyzed as *sent(assert, svoc)*, an assertive sentence consisting of a subject, a verb, an object and a complement.

[1] These are constructions we used in MINCAL. In addition in various experiments we have used a few dozen other constructions, e.g. those covering "open idioms" (see Section 4).

[2] As usual, current situations are described on the object level, and the metalevel is a place for rules that can eliminate some of the models permitted by the object level and the referential level.

Google Exhibit 1020
Page 0076

The latter and the message of the imperative that is passed to *sent(assert, svoc)* does not contain any structural information about the attachment of the PPs. This message is combined with the messages of the verb and the noun, yielding

```
[  den want(other_agent)]
 [  agent hearer]
 [  mental_agent
   [  [  type person]
      [  den speaker]
[  [  action
      [  [  den arrange]
   [  action_object
         [  type event]
         [  den conference]
         [  number 1]
         [  mods
            [  [  det a]
               [  pp_msg
                  [  [  prep at]
                     [  type time(hour)]
                     [  den
                        [  [  hour
                              [  5 am_or_pm]
                           [  minute 0]
                  [  [  prep in]
                     [  type place ]
                     [  den office]
                     [  mods
                        [  [  det my]
```

This result of parsing is then interpreted by the domain interpreter to produce:

```
***Slots:
[  [  action_name schedule]
   [  event_name
      [  a conference]
   [  event_time
      [  [  minute 0]
         [  hour
            [  5 am_or_pm]
   [  event_place
      [  my office]
```

Application-specific defaults then produce yet another interpretation where, in addition to filling the slots of Xdiary, [hour [5 am_or_pm]] is interpreted as [hour [17]].

The parser is a chart parser, working left to right, with no lookahead. The grammar is L-attributed, i.e., has has both synthesized and inherited attributes, but each inherited attribute depends only on inherited attributes of the parent or attributes of the sisters to the left. Hence, although the parser does not have a lookahead step at present, such a step can be added following [2].

4 Comparisons with related work

Linguistic arguments for constructions-based grammars has been worked out chiefly by Ch. Fillmore and his colleagues (cf. [3]). Their motivation for advocating such an approach comes from the fact that typical generative theories of grammar cannot deal properly with open idioms illustrated by constructions such as:

The more carefully you work, the easier it will get.

Why not fix it yourself?

Much as I like Ronnie, I don't approve of anything he does.

It's time you brushed your teeth.

Him be a doctor?

The same is true about even so-called robust parsers of English. The reason for this failure can be attributed to the fact that expressions like these "exhibit properties that are not fully predictable from independently known properties of its lexical make-up and its grammatical structure" – [3], p.511. However we do not need a list of "strange" constructions to conclude that thoroughly integrating syntax with semantics and pragmatics could provide us with a better handle on natural language understanding. On a closer examination "normal" constructions exhibit enough complexity to warrant the new approach (see [10] for details).

Jurafsky [4] has independently come up with a proposal for a computable grammar of constructions. We compare our work with his in [10]. Here, we limit ourselves to a few remarks. What is common in both approaches is the centrality of the concept of grammatical construction as a data structure that represents lexical, semantic and syntactic knowledge. However, there are important differences between the two formalisms. First, the actual data structures used to represent constructions are different. The most important difference has to do with the presence of the *context* field in our version of the construction grammar. This allows us to account for the importance of pragmatics in representing many constructions, and to deal with discourse constructions.

Secondly, while Jurafsky acknowledges the need for abstract constructions (pp.43-51), his abstract constructions (*weak constructions*) are not first class citizens — they are defined only extensionally, by specifying the set of constructions they abstract over, and their abstract meaning (e.g. **entity** for NOUN). They are used to simplify descriptions of constituents of other constructions. However, because they do not have a separate *vehicle* part, they cannot be used to assign default meanings. For instance, since *verb* is defined as a collection of all verbs *is + read + cancel + know + look-up + ...*, it cannot be assigned a feature *action_verb* without introducing a contra-

Google Exhibit 1020
Page 0077

diction – its semantics is therefore given as RELATION/PROCESS. For us the important feature of "abstract" constructions is not that they simplify descriptions of other constructions, but that they have default meanings. (A similar critique of [5] can be found in [10]).

5 Summary of results

Our approach to NLU is based both on linguistic arguments and on our dissatisfaction with the state of the art. State of the art systems typically are too "syntax-driven", failing to take context into account in determining the intended meaning of sentences. A related further weakness is that such systems are typically "sentence oriented", rather than "conversation/discourse oriented". In our view, this makes even the most robust systems "brittle" and ultimately impractical.

To test whether a construction-based approach is feasible built a "complete" working system that would include a representation for constructions. To do this, we focused on the "calendar domain", a domain with limited complexity and simple but not uninteresting semantics. We have chosen to deal with simple actions, and not e.g. with question answering, where deeper understanding would be necessary. [3]

Our contributions:
1. We have proposed *a new kind of grammar* – computable construction grammars, which are neither semantic, nor syntactic. Instead, their "productions" combine lexical, syntactic, semantic and pragmatic information. [4]
2. We have described *data structures for constructions*, and have shown that they can be effectively used by the parser. Note that the same data structure is used to encode the lexicon and the "syntactic" forms.
3. We have shown *how to parse with constructions*. We have implemented a simple chart parsing algorithm, which can be easily extended to an Eearly-like parser, as long as the construction grammar remains L-attributed. We have found that even a simple parser of construction can be quite efficient. This is partly due to the fact that it does not require copying of all syntactic and semantic information from daughters to mothers; the goal of parsing consists in producing an interpretation, and structural information can be discarded once an interpretation of a phrase is produced. It is also worth emphasizing that invoking domain semantics drastically reduces the number of parses constructed.
4. We have proposed *a modular architecture for NL interfaces* based on the division between linguistic knowledge, domain knowledge base, and application knowledge base. Based on our experience, we believe that this architecture should work in general for speech-enabled interfaces for restricted domains.

[3] We have also thought about another possibility, that is, enhancing an IR system, e.g. with the understanding of date expressions.

[4] In what sense are they "computable"? Although this adjective might suggest a formal model with computational complexity results, etc., what we have in mind is pretty trivial: (1) the system actually computes the messages of grammatical construction; (2) the grammars and constructions are well defined data structures, and parsing (combining all associated constructions in all possible ways) is decidable.

References

[1] *Computational Linguistics.* Special Issue on the Lexicon, 1987. 13: 3-4.

[2] Nelson Correa. An extension of earley's algorithm for s- and l- attributed grammars. In *Proc. Intl. Conf On Current Issues in Computational Linguistics*, pages 248–261. Penang, Malaysia, 1991.

[3] Charles J. Fillmore, Paul Kay, and Mary Catherine O'Connor. Regularity and idiomaticity in grammatical constructions. *Language*, 64(3):501–538, 1988.

[4] D. Jurafsky. *An On-line Computational Model of Sentence Interpretation.* PhD thesis, University of California, Berkeley, 1992. Report No. UCB/CSD 92/676.

[5] Paul Kay. Even. *Linguistics and Philosophy*, 13(1):59–111, 1990.

[6] M. McCord, Arendse Bernth, Shalom Lappin, and Wlodek Zadrozny. Natural language processing technology based on slot grammar. *International Journal on Artificial Intelligence Tools*, 1(2):229–297, 1992.

[7] I. Melcuk. *Dependency Syntax: Theory and Practice.* State University of New York Press, Albany, NY, 1988.

[8] Wlodek Zadrozny. Reasoning with background knowledge – a three-level theory. *Computational Intelligence*, 10(2), 1994. To appear.

[9] Wlodek Zadrozny and Karen Jensen. Semantics of paragraphs. *Computational Linguistics*, 17(2):171–210, 1991.

[10] Wlodek Zadrozny and Alexis Manaster-Ramer. The significance of constructions. *Submitted to Computational Linguistics*, 1994.

©COLING 94 Organizing Committee
All rights reserved. No part of this publication may be reproduced, stored in retrieval system or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior notice.

無断転載を禁ず。

3 1198 02846 8374
N/1198/02846/8374X

Appendix 1006C

https://franklin.library.upenn.edu/catalog/FRANKLIN_9919010503503681

Penn Libraries
University of Pennsylvania | Franklin

Keyword | Find books, journals, videos, & more | Advanced | Bookmarks (0)

Everything | Catalog | Articles+ | Databases | Colenda | Website

Back to Search Results | Start Over

COLING 94 : the 15th International Conference on Computational Linguistics : proceedings : August 5-9, 1994, Kyoto, Japan.

Publication: [Kyoto, Japan] : COLING 94 Organizing Comm., [1994]

Format/Description: Conference/Event
Book
2 volumes (1299 pages) : illustrations ; 26 cm

Conference Name: International Conference on Computational Linguistics (15th : 1994 : Kyoto, Japan)

Status/Location: **Available** LIBRA P98 .I55 1994 - v.1-2 -

Tools

- Bookmark
- Login to Email
- Cite
- Print
- Download in RIS format
- Staff View

Get It

Select an Item

Click here to see all Items

The Penn Libraries is committed to describing library materials using accurate, inclusive, and non-derogatory language. For more information, see our Statement on Harmful Language

Select an Item

Click here to see all Items

Login to Activate Requesting Options

describing library materials using accurate, inclusive, and non-derogatory language. For more information, see our Statement on Harmful Language

Details

Other Title:	COLING94.
Subjects:	Computational linguistics -- Congresses. Computational linguistics.
Form/Genre:	Conference papers and proceedings.
Contents:	v.1. Machine translation ; Morphology & tagging ; Lexicon ; Generation ; Parsing ; Computational linguistics ; Corpus-based NLP v.2. Semantics ; Syntax ; Speech & phonology ; Information retrieval & extraction ; Discourse & pragmatics ; Reserve paper[s].
Notes:	Includes bibliographical references and indexes.
OCLC:	80479767

PENN LIBRARIES
Library Home
Libraries & Hours
BorrowDirect+/ILL
Make an appointment

FRANKLIN HELP
Your Library Account
Using Franklin Catalog
Using Franklin Articles+
Accessing E-Resources

CONTACT US
Questions or comments

Ask

© University of Pennsylvania

Staff View

LEADER 01702cam a2200349 a 4500

005			20220608195453.0
008			950622s1994 ja a b 101 0 eng d
001			9919010503503681
035			**a**l (OCoLC)ocm80479767
035			**a**l (OCoLC)80479767
035			**a**l (CStRLIN)PAUGEBH7914-B
035			**9**l EBH7914
035			**a**l 1901050
035			**a**l (PU)1901050-penndb-Voyager
040			**a**l PU **b**l eng **c**l PU
050		4	**a**l P98 **b**l .I55 1994
111	2		**a**l International Conference on Computational Linguistics **n**l (15th : **d**l 1994 : **c**l Kyoto, Japan)
245	1	0	**a**l COLING 94 : **b**l the 15th International Conference on Computational Linguistics : proceedings : August 5-9, 1994, Kyoto, Japan.
264		1	**a**l [Kyoto, Japan] : **b**l COLING 94 Organizing Comm., **c**l [1994]
300			**a**l 2 volumes (1299 pages) : **b**l illustrations ; **c**l 26 cm
336			**a**l text **b**l txt **2**l rdacontent
337			**a**l unmediated **b**l n **2**l rdamedia
338			**a**l volume **b**l nc **2**l rdacarrier
504			**a**l Includes bibliographical references and indexes.
505	0		**a**l v.1. Machine translation ; Morphology & tagging ; Lexicon ; Generation ; Parsing ; Computational linguistics ; Corpus-based NLP -- v.2. Semantics ; Syntax ; Speech & phonology ; Information retrieval & extraction ; Discourse & pragmatics ; Reserve paper[s].
650		0	**a**l Computational linguistics **v**l Congresses. **0**l http://id.loc.gov/authorities/subjects/sh2008101459
650		7	**a**l Computational linguistics. **2**l fast **0**l http://id.worldcat.org/fast/871998
655		7	**a**l Conference papers and proceedings. **2**l lcgft **0**l http://id.loc.gov/authorities/genreForms/gf2014026068

655	7	**a|** Conference papers and proceedings. **2|** fast **0|** http://id.worldcat.org/fast/1423772
740	0	**a|** COLING94.
902		**a|** MARCIVE 2022
948		**a|** ajs#

Appendix 1006D

https://hollis.harvard.edu/primo-explore/fulldisplay?docid=01HVD_ALMA211974327780003941&context=L&vid=HVD2&lang=en

BOOK

COLING 94 : the 15th International Conference on Computational Linguistics, August 5-9, 1994, Kyoto, Japan

International Conference on Computational Linguistics (15th : 1994 : Kyoto) / [S.l.] : ICCL
c1994

WIDENER LIBRARY WID-LC (P98 .I55 1994X)

MAP IT: WIDENER 2ND FLOOR WEST, ROW 56

Send to

TEXT CALL # | EXPORT TO EXCEL | PERMALINK | E-MAIL | CITATION | EXPORT RIS | EXPORT BIBTEX | PRINT

RESERVES LIST | REPORT A PROBLEM

https://id.lib.harvard.edu/alma/990059738670203941/catalog

COPY THE PERMALINK TO CLIPBOARD

Details

Title	COLING 94 : the 15th International Conference on Computational Linguistics, August 5-9, 1994, Kyoto, Japan
Author / Creator	International Conference on Computational Linguistics (15th : 1994 : Kyoto)
Variant title	C.O.L.I.N.G. 94.
Published	[S.l.] : ICCL, c1994.
Description	2 v. : ill. ; 26 cm.
Language	English
Notes	Title on cover - "The 15th International Conference on Computational Linguistics". Includes bibliographical references and indexes.
Subjects	Computational linguistics -- Congresses
Creation Date	c1994
HOLLIS number	990059738670203941
Permalink	https://id.lib.harvard.edu/alma/990059738670203941/catalog
Source	HVD - Core

View in WorldCat
View source (MARC)
Search HathiTrust (via Harvard membership)

Access Options

Harvard affiliates: sign in to see all request options Sign in

SCAN & DELIVER (BOOK CHAPTER)

Select a location below, or use the drop-downs to filter, then select a location.

Year All **Volume** All **Description** All

Widener Library WID-LC P98 .I55 1994x

HOURS & INFO

Hide Notes & Holdings

Holdings 2v.

1 - 2 of 2 Records

BARCODE	TYPE	POLICY	DESCRIPTION	STATUS	OPTIONS
32044021354139	Book	Please sign in	vol.2	In library or storage (0 requests)	
32044021354121	Book	Please sign in	vol.1	In library or storage (0 requests)	

https://hollis.harvard.edu/primo-explore/sourceRecord?vid=HVD2&docId=01HVD_ALMA211974327780003941

```
Leader 00846nam a2200229Ka 4500
001    990059738670203941
005    20020606163548.3
008    960222s1994    xx a     b    100 0 eng d
035    ##$a(OCoLC)34240603$0http://www.worldcat.org/oclc/34240603
040    ##$aVYR$cVYR$dHLS
090    ##$aP98$b.I55 1994
111    2#$aInternational Conference on Computational Linguistics$n(15th :$d1994 :$cKyoto)
245    10$aCOLING 94 :$bthe 15th International Conference on Computational Linguistics, August 5-9, 1994, Kyoto, Japan.
246    3#$aC.O.L.I.N.G. 94.
260    ##$a[S.l.] :$bICCL,$cc1994.
300    ##$a2 v. :$bill. ;$c26 cm.
500    ##$aTitle on cover - "The 15th International Conference on Computational Linguistics".
504    ##$aIncludes bibliographical references and indexes.
650    #0$aComputational linguistics$vCongresses.$0http://id.loc.gov/authorities/subjects/sh2008101459$2LCSH$041-LIBRARY_OF_CONGRESS-sh2008101459$9Y
900    ##$a005973867
901    2#$aBDH0342
903    ##$aMHBDH03422HU
905    ##$aP^^0098
998    ##$aNO-OVERLAY
AVA    ##$0990059738670203941$8221974327770003941$a01HVD_INST$bWID$cWID-LC$dP98 .I55 1994x$eavailable$f2$g0$jWIDLC$k0$p1$qWidener Library
INST   ##$a01HVD_INST
INT    ##$aP
```

HARVARD LIBRARY

LOCATIONS & HOURS
DATABASES
ASK A LIBRARIAN
REPORT A PROBLEM
HARMFUL LANGUAGE STATEMENTS
HOLLIS IMAGES
HOLLIS FOR ARCHIVAL DISCOVERY
DIGITAL COLLECTIONS
CURIOSITY

Accessibility Privacy

Creative Commons Attribution 4.0 International License

Except where otherwise noted, this work is subject to a Creative Commons Attribution 4.0 International License, which allows anyone to share and adapt our material as long as proper attribu is given. For details and exceptions, see the Harvard Library Copyright Policy ©2021 Presidents and Fellows of Harvard Colleg

Appendix 1006E

https://web.archive.org/web/19980110012724/http://www.franklin.library.upenn.edu/ 150%

http://www.franklin.library.upenn.edu/ Go

917 captures
10 Jan 1998 – 1 Apr 2024

DEC JAN APR
1997 10 1998 1999

UNIVERSITY OF PENNSYLVANIA
L·I·B·R·A·R·Y

Franklin/Web, Penn's Online Catalog

Franklin/Web, Penn's Online Catalog has moved to http://www.franklin.library.upenn.edu/.

Please adjust your links and bookmarks.

---

https://web.archive.org/web/19971210172035/http://hplus.harvard.edu/alpha/HOLLIS.html 150%

http://hplus.harvard.edu/alpha/HOLLIS.html Go

16 captures
10 Dec 1997 - 9 Jun 2007

NOV DEC MAY
1996 10 1997 1999

HOLLIS (Harvard OnLine Library Information System)

Abstract: HOLLIS (Harvard OnLine Library Information System) is composed of 13 individual databases which include the holdings of the Harv University libraries, a course reserves database, a guide to Harvard libraries, and 9 citation indexes covering a variety of subject areas.

More Details: About HOLLIS (Harvard OnLine Library Information System)

Subjects: Library Catalogs

Connect to HOLLIS (Harvard OnLine Library Information System)

HOLLIS Plus Home | **News** | **Search** | **Comments** | **Help**

Appendix 1006F

Appendix 1006F
Citations to *COLING 94*

Source Document

COLING 94 : the 15th International Conference on Computational Linguistics : Proceedings : August 5-9, 1994, Kyoto, Japan

Citations to articles in *COLING 94*: 1995 to 1998 (in chronological order)

Rigau, G., Rodríguez, H., & Turmo, J. (1995). Automatically extracting Translation Links using a wide coverage semantic taxonomy. In *proceedings fifteenth International Conference AI 95, Language Engineering* (Vol. 95).

This article cites Ageno, A., Ribas, F., Rigau, G., Rodriguez, H., & Samiotou, A. (1994). TGE: Tlinks Generation Environment. In *COLING 94 :* the *15th International Conference on Computational Linguistics : Proceedings*.

Mikheev, A. (1996). *Domain knowledge for natural language processing*. HCRC Publications, University of Edinburgh.

This article cites Zadrozny, W., Szummer, M., Jarecki, S., Johnson, D. E., & Morgenstern, L. (1994). NL Understanding with a Grammar of Constructions. In *COLING 94 : the 15th International Conference on Computational Linguistics : Proceedings*.

Grammar, I., & Tapanainen, P. (1997). 10 Applying a Finite-State. *Finite-state Language Processing*, 311.

This article cites Karttunen, L. (1994). Constructing Lexical Transducers. In *COLING 94 : the 15th International Conference on Computational Linguistics : Proceedings*.

Yoon, J., Kim, S., & Song, M. (1997, August). Ambiguity Resolution Using Lexical Association. In *ROCLING 1997 Poster Papers* (pp. 267-271).

This article cites Kobayasi, Y., Takenobu, T., & Hozumi, T. (1994). Analysis of Japanese Compound Nouns Using Collocational Information. In *COLING 94 : the 15th International Conference on Computational Linguistics : Proceedings*.

Claramunt, G. R., & Hontoria, H. R. (1998). *Automatic acquisition of lexical knowledge from MRDs* (Doctoral dissertation, PhD Thesis. Departament de Llenguatges i Sistemes Informàtics, Barcelona).

This article cites Ageno, A., Ribas, F., Rigau, G., Rodriguez, H., & Samiotou, A. (1994). TGE: Tlinks Generation Environment. In *COLING 94 :* the *15th International Conference on Computational Linguistics : Proceedings*.

Appendix 1008A

LIBRARY | CATALOG

1 of 1

BOOK

Handbook of multimodal and spoken dialogue systems : resources, terminology, ...

Full Record MARC Tags

Main title

Handbook of multimodal and spoken dialogue systems : resources, terminology, and product evaluation / edited by Dafydd Gibbon, Inge Mertins, Roger K. Moore.

Published/Created

Boston, MA : Kluwer Academic Publishers, c2000.

Request this Item | LC Find It

More Information

LCCN Permalink	https://lccn.loc.gov/00044780
Description	xviii, 519 p. : ill. ; 25 cm.
ISBN	0792379047 (alk. paper)
LC classification	QA76.9.N38 H362 2000
Related names	Gibbon, Dafydd. Mertins, Inge, 1964- Moore, Roger, 1952-

LC Subjects	Natural language processing (Computer science) Automatic speech recognition.
Browse by shelf order	QA76.9.N38
Notes	Includes bibliographical references (p. 330-358) and index.
Series	The Kluwer international series in engineering and computer science ; SECS 565
LCCN	00044780
Dewey class no.	006.3/5
Type of material	Book

Item Availability

CALL NUMBER	QA76.9.N38 H362 2000 FT MEADE Copy 1
Request in	Jefferson or Adams Building Reading Rooms - STORED OFFSITE
Status	Not Charged

CALL NUMBER	QA76.9.N38 H362 2000 FT MEADE Copy 2
Request in	Jefferson or Adams Building Reading Rooms - STORED OFFSITE
Status	Not Charged

LIBRARY | CATALOG

1 of 1

BOOK

Handbook of multimodal and spoken dialogue systems : resources, terminology, ...

Full Record **MARC Tags**

000		01316cam a22002894a 4500
001		12032900
005		20001030102513.0
008		000607s2000 maua b 001 0 eng
906	__	|**a** 7 |**b** cbc |**c** orignew |**d** 1 |**e** ocip |**f** 20 |**g** y-gencatlg
925	0_	|**a** acquire |**b** 2 shelf copies |**x** policy default
955	__	|**a** to ASCD pc21 06-07-00; jf03 06-09-00 ; jf11 to sl 6-12-00; jf12 to Dewey 06-12-00; aa05 06-13-00; jf00 09-14-00; CIP ver. jf07 to SL 10-03-00; jf12 to BCCD 10-13-00; copy 2 added jf16 to BCCD 10-30-00
010	__	|**a** 00044780
020	__	|**a** 0792379047 (alk. paper)
040	__	|**a** DLC |**c** DLC
042	__	|**a** pcc
050	00	|**a** QA76.9.N38 |**b** H362 2000
082	00	|**a** 006.3/5 |**2** 21
245	00	|**a** Handbook of multimodal and spoken dialogue systems : |**b** resources, terminology, and product evaluation / |**c** edited by Dafydd Gibbon, Inge Mertins, Roger K. Moore.
260	__	|**a** Boston, MA : |**b** Kluwer Academic Publishers, |**c** c2000.
300	__	|**a** xviii, 519 p. : |**b** ill. ; |**c** 25 cm.
440	_4	|**a** The Kluwer international series in engineering and computer science ; |**v** SECS 565
504	__	|**a** Includes bibliographical references (p. 330-358) and index.
650	_0	|**a** Natural language processing (Computer science)

650	_0	|**a** Automatic speech recognition.
700	1_	|**a** Gibbon, Dafydd.
700	1_	|**a** Mertins, Inge, |**d** 1964-
700	1_	|**a** Moore, Roger, |**d** 1952-

Request this Item

LC Find It

Item Availability

CALL NUMBER	QA76.9.N38 H362 2000 FT MEADE Copy 1
Request in	Jefferson or Adams Building Reading Rooms - STORED OFFSITE
Status	Not Charged

CALL NUMBER	QA76.9.N38 H362 2000 FT MEADE Copy 2
Request in	Jefferson or Adams Building Reading Rooms - STORED OFFSITE
Status	Not Charged

Appendix 1008B

Copyright United States Copyright Office

Try the Copyright Public Records System (CPRS) pilot with enhanced search features and filters.

Help | Search | History | Titles | Start Over

Public Catalog

Copyright Catalog (1978 to present) at DC4
Search Request: Left Anchored Title = Handbook of multimodal and spoken dialogue systems
Search Results: Displaying 1 of 1 entries

previous next

Labeled View

Handbook of multimodal and spoken dialogue systems : resources,...

Type of Work: Text
Registration Number / Date: TX0005280987 / 2000-09-19
Title: Handbook of multimodal and spoken dialogue systems : resources, temrinology, and product evaluation / edited by Dafydd Gibbon, Inge Mertins, Roger K. Moore.
Imprint: Boston : Kluwer Academic Publishers, c2000.
Description: 519 p.
Series: The Kluwer international series in engineering and computer science ; no. SECS 565
Copyright Claimant: Kluwer Academic Publishers
Date of Creation: 2000
Date of Publication: 2000-08-14
Other Title: The Kluwer international series in engineering and computer science ; no. SECS 565
Names: Gibbon, Dafydd
Mertins, Inge
Moore, Roger K.
Kluwer Academic Publishers

previous next

Save, Print and Email (Help Page)		
Select Download Format	Full Record	Format for Print/Save
Enter your email address:		Email

Help Search History Titles Start Over

Contact Us | Request Copies | Get a Search Estimate | Frequently Asked Questions (FAQs) about Copyright | Copyright Office Home Page | Library of Congress Home Page

Google Exhibit 1020
Page 00102

Appendix 1008C

https://web.archive.org/web/20000229084601/http://catalog.loc.gov/ 133%

http://catalog.loc.gov/ Go OCT FEB MAR

5,797 captures

12 Oct 1999 – 2 Apr 2

1999 29 2000 2001

The LIBRARY of CONGRESS

Online Catalog

Frequently Asked Questions

About This Catalog | Searching Help | Services for Researchers | Search Other Libraries' Catalogs

Select a Search Method	Description *(consult **Search Examples** within each search method for more information)*
Subj-Name-Title-Call#	Browse the catalog by subject, name (personal, corporate or meeting), or call number; search by title or serial title. Search limits are available only for title and serial title searches.
Guided Keyword	Use a "fill-in" form to create searches. Search words or phrases may be restricted to particular indexes and combined using Boolean operators. Number searches and search limits are available.
Command Keyword	Use Boolean operators and/or index codes in a command mode. Number searches and search limits are available.
Keyword	Search for words or exact phrases anywhere in the catalog record. Add punctuation for more precise results (see search examples). Results are relevance ranked. Search limits are available.

The Wayback Machine - https://web.archive.org/web/20000229084601/http://catalog.loc.gov:80/

The LIBRARY of CONGRESS Online Catalog

Frequently Asked Questions

About This Catalog | Searching Help | Services for Researchers | Search Other Libraries' Catalogs

Select a Search Method	**Description** *(consult **Search Examples** within each search method for more information)*
Subj-Name-Title-Call#	Browse the catalog by subject, name (personal, corporate or meeting), or call number; search by title or serial title. Search limits are available only for title and serial title searches.
Guided Keyword	Use a "fill-in" form to create searches. Search words or phrases may be restricted to particular indexes and combined using Boolean operators. Number searches and search limits are available.
Command Keyword	Use Boolean operators and/or index codes in a command mode. Number searches and search limits are available.
Keyword	Search for words or exact phrases anywhere in the catalog record. Add punctuation for more precise results (see search examples). Results are relevance ranked. Search limits are available.

Set Search Limits	Limits restrict search results by language, date, type of material (e.g., book, music, etc.), location of material in the Library, or place of publication. Set search limits now, or prior to executing a search; they remain in effect until cleared or changed.

Information about the images: Two pendentive paintings by Edward J. Holslag are displayed from the Librarian's Room (Librarian's Ceremonial Office) located in the Thomas Jefferson Building of the Library of Congress. On the left, "Efficiunt clarum studio" (Study, the watchword of fame); on the right, "Dulce ante omnia musae" (The Muses, above all things, delightful).

Library of Congress Home | Using the Library | Thomas | Copyright Office | American Memory | Exhibitions | The Library Today | Bicentennial | Help & FAQs | Search our Site | Site Map

Library of Congress

Comments: lcweb@loc.gov (02/11/2000)

Google Exhibit 1020
Page 00105

Appendix 1008D

Appendix 1008D
Citations to Gibbon

Source Document

Handbook of Multimodal and Spoken Dialogue Systems : Resources, Terminology, and Product Evaluation. Edited by Dafydd Gibbon, Inge Mertins, Roger K. Moore. Boston, MA : Kluwer Academic Publishers, 2000.

Citations: 2001 to 2002 (in chronological order)

Weisser, M. (2001, May). *A Corpus-Based Methodology for Comparing and Evaluating Native and Non-Native Speaker Accents* (Doctoral dissertation, University of Lancaster).

Kurematsu, A., Hosoki, M., & Morimoto, Y. (2001, October). Prosody Pattern Recognition and Identification of Utterance Intentions in Japanese Spontaneous Speech. In *ISCA Tutorial and Research Workshop (ITRW) on Prosody in Speech Recognition and Understanding*.

Krahmer, E. J. (2001). The Science and Art of Voice Interfaces. (Philips Research Report). Philips.

Lemon, O., Bracy, A., Gruenstein, A., & Peters, S. (2001). Information states in a multi-modal dialogue system for human-robot conversation. In *5th Workshop on Formal Semantics and Pragmatics of Dialogue (Bi-Dialog 2001)* (pp. 57-67).

Suhm, B., & Peterson, P. (2002, January). A data-driven methodology for evaluating and optimizing call center IVRs. *International Journal of Speech Technology*, *5*, 23-37.

Appendix 1016A

https://search.library.wisc.edu/catalog/999805139302121

UNIVERSITY OF WISCONSIN–MADISON LIBRARIES

UW SIGN IN HELP GIVE

LIBRARIES SEARCH

ARTICLES CATALOG DATABASES JOURNALS UWDC Ask

CATALOG SEARCH

Saved Lists

Keywords

Books, music, movies, and more

Search > Catalog > Industrial parsing of software manuals

Books

Industrial parsing of software manuals

AVAILABLE AS Physical

Google Preview

Request a Copy

Save to Lists

Citation

Details

Physical Locations

Memorial Library **Stacks Regular Size Shelving**

PE1074.5 I48 1996

https://search.library.wisc.edu/catalog/999805139302121/staff

Subjects

Linguistic String Parser (Computer grammar)

Details

Physical Locations

Memorial Library
Today's hours: 10 a.m.–9 p.m.

Stacks Regular Size Shelving
PE1074.5 I48 1996

Publication Details

CREATOR	editors, Richard F.E. Sutcliffe, Heinz-Detlev Koch and Annette McElligott
FORMAT	Books
LANGUAGE	English
CONTRIBUTORS	Sutcliffe, Richard F. E., 1960- Koch, Heinz-Detlev McElligott, Annette
PUBLICATION	Amsterdam ; Atlanta, GA : Rodopi, 1996
PHYSICAL DETAILS	277 pages : illustrations ; 23 cm
ISBNS	9789042001145, 9042001143
OCLC	ocm36467363

Notes

Includes bibliographical references and index.

Subjects

Linguistic String Parser (Computer grammar)

Library Staff Details

MARC Bibliographic Record

LEADER		00989mam a2200289 a 4500
001		9938017643602122
005		20180322223632.0
008		970228s1996 ne a b 001 0 eng d
020		$a9042001143
035		$a(OCoLC)36467363
035		$9AVU5116UW
035		$a(WU)3801764-uwmadisondb
035		$a(EXLNZ-01UWI_NETWORK)999805139302121
040		$aNYP$beng$cNYP$dCUY$dGZM
049		$aGZMA
090		$aPE1074.5$b.I48 1996
245	00	$aIndustrial parsing of software manuals /$ceditors, Richard F.E. Sutcliffe, Heinz-Detlev Koch and Annette McElligott.
264	_1	$aAmsterdam ;$aAtlanta, GA :$bRodopi,$c1996.
300		$a277 pages :$billustrations ;$c23 cm

336		$atext$btxt$2rdacontent
337		$aunmediated$bn$2rdamedia
338		$avolume$bnc$2rdacarrier
504		$aIncludes bibliographical references and index.
650	_0	$aLinguistic String Parser (Computer grammar)
700	1_	$aSutcliffe, Richard F. E.,$d1960-
700	1_	$aKoch, Heinz-Detlev.
700	1_	$aMcElligott, Annette.
997		$aMARCIVE

MMS IDs

Document ID: 999805139302121
Network Electronic IDs:
Network Physical IDs: 999805139302121
mms_mad_ids: 9938017643602122

WISCONSIN

Libraries

Accessibility

Copyright Statement

Policies

Contact Us

Support Libraries

Appendix 1016B

BOOK

Industrial parsing of software manuals / editors, Richard F.E. Sutcliffe, Heinz-Detlev Koch and Annette McElligott.

Sutcliffe, Richard F. E.; Koch, Heinz-Detlev.; McElligott, Annette.

1996

Available at Doe Library Main (Gardner) Stacks (PE1074.5 .I48 1996)

TOP

GET IT

LINKS

DETAILS

SEND TO

VIRTUAL BROWSE

Get It

Sign in to view request options and to access your account. SIGN IN

BACK TO LOCATIONS

LOCATION ITEMS

Doe Library

Available , Main (Gardner) Stacks ; PE1074.5 .I48 1996

(1 copy, 1 available, 0 requests)

Item in place

Loanable

OTHER UC LIBRARIES

University of California Irvine

Available in institution

University of California Los Angeles

Available in institution

Links

Display MARC Record

Details

Title	Industrial parsing of software manuals / editors, Richard F.E. Sutcliffe, Heinz-Detlev Koch and Annette McElligott.
Format	277 pages : illustrations ; 23 cm
Creation Date	1996
Publisher	Amsterdam ; Atlanta, GA : Rodopi, 1996.
Series	Language and computers: studies in practical linguistics ; no. 17
Notes	Includes bibliographical references and index.
Terms governing use and reproduction	Current Copyright Fee: GBP20.00 0.
Subject	Lingustic String Parser (Computer grammar) Handleidingen Parsing Software Datenverarbeitungssystem Syntaktische Analyse Parser
Contributor	Sutcliffe, Richard F. E. Koch, Heinz-Detlev. McElligott, Annette.
Identifier	ISBN : 9042001143 ISBN : 9789042001145 ISBN : 904200102X ISBN : 9789042001022
Source	Library Catalog

```
leader  01917cam a2200565 a 4500
001     9913109079506531
005     20230504213020.0
008     970228s1996    ne a     b    001 0 eng d
016     7#$a006374558 $2Uk
016     7#$aB9605469 $2bccb
020     ##$a9042001143
020     ##$a9789042001145
020     ##$a904200102X
020     ##$a9789042001022
035     ##$a(CLU)4299249-ucladb-Voyager
035     ##$a(OCoLC)36467363
035     ##$a(OCoLC)ocm36467363
040     ##$aNYP $beng $cNYP $dCUY $dOCL $dNLGGC $dYDXCP $dUPM $dHEBIS $dOCLCQ $dUKMGB $dOCLCO $dOCLCQ
049     ##$aMAIN
050     14$aPE1074.5 $b.I48 1996
082     04$a006.35 $qOCoLC
084     ##$a17.46 $2bcl
090     ##$aPE1074.5 $b.I48 1996 $9LOCAL
245     00$aIndustrial parsing of software manuals / $ceditors, Richard F.E. Sutcliffe, Heinz-Detlev Koch and Annette McElligott.
260     ##$aAmsterdam ; $aAtlanta, GA : $bRodopi, $c1996.
300     ##$a277 pages : $billustrations ; $c23 cm
336     ##$atext $btxt $2rdacontent
337     ##$aunmediated $bn $2rdamedia
338     ##$avolume $bnc $2rdacarrier
490     0#$aLanguage and computers: studies in practical linguistics ; $vno. 17
540     ##$aCurrent Copyright Fee: GBP20.00 $c0. $5Uk
504     ##$aIncludes bibliographical references and index.
650     #0$aLingustic String Parser (Computer grammar)
650     17$aHandleidingen. $2gtt
650     17$aParsing. $2gtt
650     17$aSoftware. $2gtt
650     #7$aDatenverarbeitungssystem $2gnd
650     #7$aSyntaktische Analyse $2gnd
650     #7$aParser $2gnd
```

```
700     1#$aSutcliffe, Richard F. E.
700     1#$aKoch, Heinz-Detlev.
700     1#$aMcElligott, Annette.
908     ##$aWorldCat Daily Updates 2022-09-19 $bWorldCat record variable field(s) change: 040, 490, LDR
908     ##$aWorldCat Daily Updates 2023-05-04 $bWorldCat record variable field(s) change: 082
968     ##$aReclvl: f Addate: 970409 Addid: OCL Moddate: 080604 Modid: XRF $9LOCAL
996     ##$eGLADN84436128 $9LOCAL
996     ##$ab142055542 $bUpdated: 08-12-11 $cCreated: 03-28-09 $dCatDate: 04-09-97 $9LOCAL
```

PRIVACY | ACCESSIBILITY | NONDISCRIMINATION

Berkeley
UNIVERSITY OF CALIFORNIA

Appendix 1016C

https://web.archive.org/web/19970607084855/http://www.lib.berkeley.edu/Catalogs/ 150%

http://www.lib.berkeley.edu/Catalogs/ Go

746 captures
7 Jun 1997 - 22 Aug 2023

MAY JUN JUL
1996 07 1997 1999

Library Online Catalog Systems

	Web-Based Systems		Text (Telnet) Mode (about Telnet)	
UC Berkeley	UCB Library Pathfinder		GLADIS	
	Connect	More Info	Connect	More Inf
UC Systemwide	MELVYL® system: Web version		MELVYL® system: Telnet version	
	Connect	More Info	Connect	More In

Links to Other Online Catalogs

Guides to GLADIS, Melvyl and other Catalogs

The Wayback Machine - https://web.archive.org/web/19970607084855/http://www.lib.berkeley.edu:80/Catalogs/

Library Online Catalog Systems

	Web-Based Systems		Text (Telnet) Mode (about Telnet)	
UC Berkeley	UCB Library Pathfinder		GLADIS	
	Connect	More Info	Connect	More Info
UC Systemwide	MELVYL® system: Web version		MELVYL® system: Telnet version	
	Connect	More Info	Connect	More Info

Links to Other Online Catalogs

Guides to GLADIS, Melvyl and other Catalogs

- GLADIS and Melvyl General Search Guides
- Special Features in the Melvyl System Catalog
- International Catalogs/Databases available via Melvyl
- Guides to catalogs from other UC campuses
- Guides to other California library catalogs
- Guides to other library catalogs in the US
- Guides to catalogs from other countries

More Help!

Library Information Guides

Library Instruction & Tours

Library Services

Copyright (C) 1997 by the Library, University of California, Berkeley. ***All rights reserved.***
*** MELVYL is a registered trademark of The Regents of the University of California**
Document maintained on server: http://www.lib.berkeley.edu/ by the Library Web Manager
Last update 5/6/97. Server manager: webman@library.berkeley.edu

Appendix 1016D

Appendix 1016D
Citations to Sutcliffe

Source Document

Sutcliffe, R. F., Koch, H. D., & McElligott, A. (Eds.). (1996). *Industrial Parsing of Software Manuals*. Rodopi.

Citations: 1997 – 2001 (in chronological order)

Churcher, G. E., Atwell, E., & Souter, C. (1997, September) A generic template for the evaluation of dialogue management systems. In: Kokkinakis, G, Fakotakis, N and Dermatas, E, (eds.) *Proceedings of EUROSPEECH 1997. EUROSPEECH 1997, 22-25 Sep 1997, Rhodes, Greece, ISCA*, 9-16.

Lin, D. (1998, June). A dependency-based method for evaluating broad-coverage parsers. *Natural Language Engineering*, *4*(2), 97-114.

Atwell, E. S., Demetriou, G., Hughes, J., Schiffrin, A., Souter, C., & Wilcock, S. (2000). A comparative evaluation of modern English corpus grammatical annotation schemes. *ICAME Journal: International Computer Archive of Modern and Medieval English Journal*, *24*, 7-23.

Mollá, D., Schneider, G., Schwitter, R., & Hess, M. (2000). Answer extraction using a dependency grammar in ExtrAns. *TAL*, *41*(1), 127-156.

Demetriou, G., & Atwell, E. S. (2001). A domain-independent semantic tagger for the study of meaning associations in English text. In *Proceedings of the Fourth International Workshop on Computational Semantics (IWCS-4)* (pp. 67-80). Association for Computational Linguistics (ACL) Special Interest Group in Computational Semantics (SIGSEM).

Appendix A

Ingrid Hsieh-Yee

Professor Emerita
Dept. of Library and Information Science
Catholic University of America
Washington, D.C. 20064
E-mail: hsiehyee@cua.edu

Education

Ph.D. 1990. Library and Information Studies, University of Wisconsin-Madison
Minors: Sociology and Psychology

M.A. 1984. Library and Information Studies, University of Wisconsin-Madison.

M.A. 1983. Comparative Literature, University of Wisconsin-Madison.

B.A. 1980. Foreign Languages and Literature, National Taiwan University.

Work Experience

Professor, School/Dept. of Library and Information Science, Catholic University of America, 1990-2022 (Assistant Professor, 1990-1996; Associate Professor, 1997-2004; Professor 2004-2022; Professor Emerita 2022)

Co-Chair, Dept. of Library and Information Science, Catholic University of America, June 2015-August 2016.

Acting Dean, School of Library and Information Science, Catholic University of America, January 2010-June 2012.

Cataloger, Dept. of Legislative Reference Library, Annapolis, Maryland, 1989-1990.

Lecturer, School of Library and Information Studies, University of Wisconsin-Madison, 1988.

Teaching Assistant, School of Library and Information Studies, University of Wisconsin-Madison, 1986-1988.

Cataloger, Health Sciences Library, University of Wisconsin-Madison, 1984-1986.

Areas of Teaching and Research Interests

Information Organization and Access; Metadata; Cataloging & Classification; Information Architecture; Information Retrieval; Digital Collections; Social Media; Scholarly

Communication; Information Behavior; Health Informatics; Human Computer Interaction; Usability Studies

Grants & Honors

Cultural Heritage Information Management Project. IMLS grant. Amount: $498,741. Period: Aug. 2012 to July 2015. Co-PI with Dr. Youngok Choi.

D.C. Health Information Technology (HIT4): Building Capacity & Providing Access in Our Nation's Capital. Dept. of Labor H2B Training Grant. Grant amount: $4,175,500. Grant period: Nov. 2011 to Dec. 2015. Partner with the Metropolitan School of Professional Studies of the Catholic University of America, Children's National Medical Center, D.C. Department of Employment Services, Holy Cross Hospital, Howard University, Center for Urban Progress, Providence Hospital, and Sibley Memorial Hospital.

Capital Health Careers Project. Department of Labor Healthcare Sector and Other High Growth and Emerging Industries Grant. Grant amount: $4,953,999. Grant period: March 2010 – February 2013. Awarded to a group of healthcare organizations and educational institutions in Washington, D.C. Providence Health Foundation of Providence Hospital (Lead institution). Part of the grant supported the development of a Master's degree program in Information Technology with a concentration in Health Information Technology offered by the School of Library and Information Science.

The Washington D.C. School Librarians Project. IMLS grant. Grant amount: $412,660. Grant period: Aug. 2007 – June 2011. The School partnered with the District of Columbia Public Schools (DCPS) and the District of Columbia Library Association to educate and mentor school media specialists for the DCPS system. PI, Jan. 2010 to June 2011.

SIG Member of the Year, American Society for Information Science and Technology (2009).

Most Outstanding Paper of *OCLC Systems & Services* (2001).

ALISE Research Grant (2001).

Most Outstanding Paper of *OCLC Systems & Services* (2000).

Research Grant from ERIC (1999-2000).

Best Research Paper Award; Association for Library and Information Science Education (1998).

Research Grants, Catholic University of America. 1991, 1992, 1993, 1996, 1998, 1999, 2004, 2005, 2006, 2007, 2013-14.

Cooperative Faculty Research Grant, Consortium of Universities in the Washington Metropolitan Area (1993-1994).

Cooperative Research Grant, Council on Library Resources (1993-1994).

Journal of the American Society for Information Science Best Paper Award (1993).

ASIS/ISI Information Science Doctoral Dissertation Scholarship (1989).

HEA Title IIB Fellowship (Dept. of Education) (1989)

Chinese-American Librarians Association Scholarship (1987).

Beta Phi Mu (1985).

Vilas Fellowship, University of Wisconsin-Madison. 1984

Publications

Hsieh-Yee, I. (2021). Can We Trust Social Media? *Internet Reference Services Quarterly*, 1-15. Published online: 28 July, 2021. 10.1080/10875301.2021.1947433.

Bailey, T., & Hsieh-Yee, I. (2020). Combating the Sharing of False Information: History, Framework & Literacy Strategies, *Internet Reference Services Quarterly*, 24 (1-2), 9-30.

Choi, Y., and Hsieh-Yee, I. (2010). Finding Images in an OPAC: Analysis of User Queries, Subject Headings, and Description Notes. *Canadian Journal of Information and Library Science,* 34(3): 271 – 295.

Hsieh-Yee, I. (2008). Educating Cataloging Professionals in a Changing Information Environment. *Journal of Education for Library and Information Science,* 46(2): 93-106.

Vellucci, S. L., Hsieh-Yee, I., and Moen, W.E. (2007). The Metadata Education and Research Information Commons (MERIC): A Collaborative Teaching and Research Initiative. *Education for Information,* 25(3&4): 169-178.

NISO Framework for Guidelines for Building Good Digital Collections. 3rd ed. Baltimore, MD: National Information Standards Organization, 2007. Also available online: http://www.niso.org/framework/framework3.pdf (NISO Working Group members: Priscilla Caplan (chair), Grace Agnew, Murtha Baca, Tony Gill, Carl Fleischhauer, Ingrid Hsieh-Yee, Jill Koelling, and Christie Stephenson.)

Choi, Y., Hsieh-Yee, I., and Kules, B. (2007). Retrieval Effectiveness of TOC and LCSH. *Proceedings of the Joint Conference on Digital Libraries*, pp. 233-234.

Vellucci, S., and Hsieh-Yee, I. (2007). They Didn't Teach Me That in Library School! Building

a Digital Teaching Commons to Enhance Metadata Teaching, Learning and Research. *Proceedings of the National Conference of the Association of College and Research Libraries, Baltimore, MD*, pp. 26-31.

Mitchell, Vanessa, and Ingrid Hsieh-Yee. (2007). Converting Ulrich's Subject Headings to FAST Headings: A Feasibility Study. *Cataloging & Classification Quarterly,* 45(1): 59-85.

Hsieh-Yee, I., Tang, R., and Zhang, S. (2007). User Perceptions of a Federated Search System. *IEEE Technical Committee on Digital Libraries Bulletin,* Summer 3(2) (URL = http://www.ieee-tcdl.org).

Tang, R., Hsieh-Yee, I., and Zhang, S. (2007). User Perceptions of MetaLib Combined Search: An Investigation of How Users Make Sense of Federated Searching." *Internet Reference Services Quarterly,* 12(12): 211-236.

Hsieh-Yee, I., Tang, R., and Zhang, S. (2006). User Perceptions of a Federated Search System. *Proceedings of the Joint Conference on Digital Libraries*, *June 11-15, 2006, Chapel Hill*, p. 338.

Hsieh-Yee, I. (2006). *Organizing Audiovisual and Electronic Resources for Access: A Cataloging Guide*. 2nd ed. Westport, Conn.: Libraries Unlimited.

NISO A Framework of Guidance for Building Good Digital Collections. 2nd ed. Bethesda, MD: National Information Standards Organization, 2004. Framework Advisory Group: Grace Agnew, Liz Bishoff, Priscilla Caplan (Chair), Rebecca Gunther and Ingrid Hsieh-Yee.

Hsieh-Yee, I. (2004). Cataloging and Metadata Education in North American LIS Programs. *Library Resources & Technical Services,* 48(1): 59-68.

Hsieh-Yee, I. (2004). Cataloging and Metadata Education. In Gary E. Gorman (Ed.), *International Yearbook of Library and Information Management 2003: Metadata Applications and Management*, (pp.204-234). London: Facet Publishing.

Yee, P. L., Hsieh-Yee, I., Pierce, G.R., Grome, R., and Schantz, L. (2004). Self-Evaluative Intrusive Thoughts Impede Successful Searching on the Internet. *Computers in Human Behavior,* 20(1): 85-101.

Hsieh-Yee, I. (2003). Cataloging and Metadata Education: A Proposal for Preparing Cataloging Professionals of the 21st Century. A report submitted to the ALCTS-Education Task Force in response to Action Item 5.1 of the *Bibliographic control of Web Resources: A Library of Congress Action Plan*. Approved by the Association for Library Collections and Technical Services. Web version available since April 2003 at http://lcweb.loc.gov/catdir/bibcontrol/CatalogingandMetadataEducation.pdf.

Hsieh-Yee, I. (2002). Cataloging and Metadata Education: Asserting a Central Role in Information Organization. *Cataloging & Classification Quarterly* 34(½): 203-222.

Hsieh-Yee, I., and Smith, M. (2001). The CORC Experience: Survey of Founding Libraries, Part I. *OCLC Systems & Services,* 17: 133-140. (Received "The Most Outstanding Paper of OCLC Systems & Services in 2001" award.)

Hsieh-Yee, I., and Smith, M. (2001). The CORC Experience: Survey of Founding Libraries, Part II, Automated Tools and Usage. *OCLC Systems & Services,* 17: 166-177. (Received "The Most Outstanding Paper of OCLC Systems & Services in 2001" award.)

Hsieh-Yee, I. (2001). ERIC User Services: Changes and Evaluation for the Future. *Government Information Quarterly,* 18: 31-42.

Hsieh-Yee, Ingrid. (2001). Research on Web Search Behavior. *Library and Information Science Research,* 23: 167-185.

Logan, E., and Hsieh-Yee, I. (2001). Library and Information Science Education in the Nineties. *Annual Review of Information Science and Technology,* 35: 425-477.

Hsieh-Yee, I. (Ed.) (2001). *Library and Information Science Research*, 23 (2). A special issue in honor of the retirement of Douglas L. Zweizig.

Hsieh-Yee, I. (2000). *ERIC User Services: Evaluation in a Decentralized Environment*. Washington, D.C.: Dept. of Education.

Hsieh-Yee, Ingrid. (2000). *Organizing Audiovisual and Electronic Resources for Access: A Cataloging Guide*. Littleton, CO: Libraries Unlimited.

Hsieh-Yee, I. (2000). Organizing Internet Resources: Teaching Cataloging Standards and Beyond. *OCLC Systems & Services,* 16: 130-143. (Received "The Most Outstanding Paper of OCLC Systems & Services in 2000" award.)

Hsieh-Yee, I. (1998). The Retrieval Power of Selected Search Engines: How Well Do They Address General Reference Questions and Subject Questions? *Reference Librarian,* 60: 27-47.

Hsieh-Yee, I. (1998). Search Tactics of Web Users in Searching for Texts, Graphics, Known Items and Subjects: A Search Simulation Study. *Reference Librarian,* 60: 61-85. (Received the 1997 Best ALISE Research Paper Award.)

Hsieh-Yee, I. (1997). Access to OCLC and Internet Resources: LIS Educators' Views and Teaching Practices. *RQ,* 36: 569-86.

Hsieh-Yee, I. (1997). Teaching Online and CD-ROM Resources: LIS Educators' Views and Practices. *Journal of Education for Library and Information Science,* 38: 14-34.

Hsieh-Yee, I. (1996). The Cataloging Practices of Special Libraries and Their Relationship with OCLC. *Special Libraries,* 87: 10-20.

Hsieh-Yee, I. (1996). Modifying Cataloging Practice and OCLC Infrastructure for Effective Organization of Internet Resources. In *Proceedings of the OCLC Internet Cataloging Colloquium*. [Online]. Available: http://www.oclc.org/oclc/man/colloq/hsieh.htm

Hsieh-Yee, I. (1996). Student Use of Online Catalogs and Other Information Channels. *College & Research Libraries,* 57: 161-175.

Hsieh-Yee, I. (1995). Ten entries in James S. C. Hu (Ed.), *Encyclopedia of Library & Information Science*, 913, 1028-29, 1036, 1037, 1145-46, 1514, 1575, 1763-64, 2216-27, 2378-79. Taipei, Taiwan: Sino-American Publishing. (Topics include "Advanced Technology/Libraries," "Information Ethics," "Instruction on Cataloging and Classification," "Instruction on Reference Services.")

Hsieh-Yee, I. (1993). Effects of Search Experience and Subject Knowledge on Online Search Behavior: Measuring the Search Tactics of Novice and Experienced Searchers. *Journal of the American Society for Information Science,* 44: 161-174. (Received the 1993 Best JASIS Paper Award.)

Works in Progress

Hsieh-Yee, I. Actionable lessons for debunking false information: A comparative analysis of literacy lessons in communication and library and information science

Hsieh-Yee, I. Representing diverse voices and communities in cataloging and indexing: The evolution of controlled vocabularies

Hsieh-Yee, I. Information Organization and Retrieval in Practice (a monograph scheduled for completion and publication in 2023)

Presentations

Bailey, T. & Hsieh-Yee, I. (February 2020). The Phenomena of Sharing Misinformation and the Need for Information Literacy. Presented at the 2020 Bridging the Spectrum Symposium, Washington, D.C., February 14, 2020.

Hsieh-Yee, I. and Fragan-Fly, J. (May 2018) Trends, Design & Strategies for Digital Scholarship Services. Presented at the 2018 Maryland/Delaware Library Association Conference, Cambridge, MD.

Hsieh-Yee, I. (February, 2018) Research Data Management: What It Takes to Succeed. Presented at the 10th Bridging the Spectrum Symposium, Washington, D.C.

Hsieh-Yee, I. (February, 2017) *Research Data Management: New Competencies and Opportunities for Information Professionals*. Presented at the 9th Bridging the Spectrum Symposium, Washington, D.C.

Hsieh-Yee, I. and Lawton, P. (February, 2017) *Enhancing Catholic Portal Searches with User Terms and LCSH*. Presented at the 9th Bridging the Spectrum Symposium, Washington, D.C.

Hsieh-Yee, I. (2016, October) *Visualizing Data for Information*. Presented at the 2016 Virginia Library Association Conference, Hot Springs, VA.

Hsieh-Yee, I. (2016, August) *Religious Materials Toolbox for Archivists: Solutions to Problems Facing the Profession*. Presented at Archives * Records 2016, Atlanta, GA.

Hsieh-Yee, I. and Lawton, P. (2016, March) *Enhancing Retrieval of Catholic Materials with LCSH Knowledge Structure*. Presented at the 2016 Catholic Library Association Conference, San Diego, CA.

Fagan-Fry, J. and Hsieh-Yee, I. (2016, February) *Approaches to Digital Scholarship at Top Universities around the World: Scholarly Publishing in the Digital Age.* Presented at the 8th Bridging the Spectrum Symposium, Washington, D.C.

Hsieh-Yee and Fagan-Fry, J. (2016, January) *Innovative Services for Digital Scholarship at Top 100 Research Libraries of the World*. Poster presented at the 2016 Annual Conference of the Association for Library and Information Science Education, Boston, Mass.

Hsieh-Yee, I. and Lawton, P. (2015, June). *Crowdsourcing terms for CRRA portal themes.* Poster presented at the third CRRA symposium and annual meeting, Bringing the created toward the Creator: Liturgical art and design since Vatican II. Catholic Theological Union, Chicago, Illinois.

Hsieh-Yee, I. and Lawton, P. (2015, February). *Crowdsourcing terms for thematic exploration in the Catholic Portal.* Poster presented at the 7th Annual Bridging the Spectrum Symposium, Washington, D.C.

Hsieh-Yee, I., James, R., and Fagan-Fry, J. (2015, February). *Support for digital scholarship at top university libraries of the world*. Poster presented at the 7th Annual Bridging the Spectrum Symposium, Washington, D.C.

Hsieh-Yee, I., Zhang, S., Lin, K., and Cherry, S. (2015, February). *Thus said the end users: Summon experience and support for research workflows*. Poster presented at the 7th Annual Bridging the Spectrum Symposium, Washington, D.C.

Yontz, E., Hsieh-Yee, I., & Houston, S. (2015, February). *Healthy Heroes Summer Reading Club: Developing healthy youth at public libraries.* 11th Annual Jean Mills Health Symposium, Greenville, North Carolina.

Yontz, E., Hsieh-Yee, I., and Houston, S. (2015, January). *Healthy youth and libraries: A pilot study.* Association for Library & Information Science Education (ALISE) Annual Conference, Chicago, Illinois.

Hsieh-Yee, I. (2014, May). *Linking CRRA resources to portal themes via authority files*. Presented at the Catholic Research Resources Alliance 2014 Membership Meeting, Marquette, WI.

Hsieh-Yee, I. (2014, April). *Enhancing subject access to CRRA resources*. Presented at the 2014 Catholic Library Association Conference, Pittsburgh, PA.

Hsieh-Yee, I. (2014, January). *Health Information Technology Program: Educational entrepreneurship in action*. Presented at the 2014 annual Conference of the Association for Library and Information Science Education, Philadelphia, PA

Hsieh-Yee, I., Zhang, S., Lin, K., and Cherry, S. (2014, January). *Discovering information through Summon: An analysis of user search strategies and search success*. Paper presented at the 6th Bridging the Spectrum Symposium, Washington, D.C.

Hsieh-Yee, I. (2012, December). *National Digital Stewardship Alliance and SLIS at CUA: An Educational Partnership*. Paper presented at Best Practices Exchange: Acquiring, Preserving, and Providing Access to Government Information in the Digital Era, Annapolis, MD

Choi, Y. and Hsieh-Yee, I. (2010, January). *Finding Images in an OPAC: Analysis of User Queries, Subject Headings, and Description Notes*. Paper presented at 2nd Annual Bridging the Spectrum Symposium, Catholic University of America, Washington, D.C.

Hsieh-Yee, I. and Coogan, J. (2010, January). *Google Scholar vs. Academic Search Premier: What Libraries and Searchers Need to Know*. Paper presented at 2nd Annual Bridging the Spectrum Symposium, Catholic University of America, Washington, D.C.

Hsieh-Yee, I. (2009, November). *Information Science Education: An LIS School's Perspective*. Paper presented at Annual Meeting of the American Society for Information Science and Technology, Vancouver, British Columbia, Canada.

Hsieh-Yee, I., Menard, E., Ya-Ning Chen, A., Shu-Jiun Chen, S., Kalfatovic, M. R., Wisser. K. M. (2009, November). *Information Organization in Libraries, Archives and Museums: Converging Practices and Collaboration Opportunities*. Presented at Annual Meeting of the American Society for Information Science and Technology, Vancouver, British Columbia, Canada. (Organizer and moderator of this panel.)

Hsieh-Yee, I. and Coogan, J. (2009, July). *Catching up to Google Scholar: The Retrieval Power of Academic Search Premier and Google Scholar*. Poster presented at American Library Association Conference, Chicago, Illinois.

Hsieh-Yee, I., with the CUA Scholarly Communications Project Team. (2009, January). *Digital Scholarship@CUA: Developing an Institutional Repository for CUA*. Poster presented at 1st Annual Bridging the Spectrum Symposium, Catholic University of America, Washington, D.C.

Wise, M., Cylke, K., and Hsieh-Yee, I. (2009, January). *Digital Talking Books: Meeting the Needs of the Blind and the Handicapped*. Paper presented at the Bridging the Spectrum Symposium, Catholic University of America, Washington, D.C.

Hsieh-Yee, I. (2009, January). *User Expectations of MERIC*. Presented at the Information Organization Competencies for the 21st Century Discussion Session of the 2009 Conference of the Association for Library and Information Science Education, Denver, Colorado.

Choi, Y., and Hsieh-Yee, I. (2008,November). *Subject Access for Images in an OPAC.* Annual Meeting of the American Society for Information Science and Technology, Columbus, Ohio. (Also co-organized a panel on Retrieving and Using Visual Resources: Challenges and Opportunities for Research and Education.)

Hsieh-Yee, I. (2008, June). *Educating Cataloging Professionals in a Changing Information Environment*. National Taiwan University, Taipei, Taiwan.

Vellucci, S. L., Moen, W.E., Hsieh-Yee, I., Marson, B., and Wisser, K. (2008, January) *Building a Metadata Education and Research Community through MERIC (Metadata Education and Research Information Commons): Demo and Stakeholder Input*. A panel presented at the 2008 Conference of the Association for Library and Information Science Education, Philadelphia, Pennsylvania.

Hsieh-Yee, I., Choi,Y. and Kules, B. (2007, October). *Searching for Books and Images in OPAC: Effects of LCSH, TOC and Subject Domains.* A poster presented at the American Society for Information Science and Technology Annual Meeting, Milwaukee, Wisconsin.

Hsieh-Yee, I. and Coogan, J. (2007, August) *Google Scholar vs. Academic Search Premier: A Comparative Analysis.* Presented to the Faculty and Staff of the University of the District

of Columbia.

Hsieh-Yee, I. and Coogan, J. (2007, June). *Google Scholar vs. Academic Search Premier: A Comparative Analysis*. Presented to the Washington Research Library Consortium Community, Catholic University of America, Washington, D.C.

Hsieh-Yee, I., Choi, Y., and Kules, B.. (2007, June). *What Users Need for Subject Access: Table of Contents or Subject Headings?* A poster presented at the 2007 American Library Association Annual Conference, Washington, D.C., June 2007.

Choi, Y., Hsieh-Yee, I., and Kules, B. (2007, June). *Retrieval Effectiveness of TOC and LCSH.* A paper presented at the Joint Conference on Digital Libraries 2007, Vancouver, Canada.

Vellucci, S. L., Hsieh-Yee, I., and Moen, W.E. (2007, May). *If We Build It, Will They Come? Building a Community of Practice for Metadata Stakeholders.* A poster presented at the Rutgers University Research Day, Bridgeton, New Jersey.

Hsieh-Yee, I. (2007, May). *Federated Searching: User Experience & Perceptions.* International Conference on Information Organization & Retrieval, National Taiwan University, Taipei, Taiwan.

Hsieh-Yee, I. (2007, May). *Search Performance of Google Scholar and Academic Search Premier.* International Conference on Information Organization & Retrieval, National Taiwan University, Taipei, Taiwan.

Hsieh-Yee, I. (2007, May) *MERIC: Building a Digital Commons for Metadata Education & Research.* International Conference on Information Organization & Retrieval, National Taiwan University, Taipei, Taiwan.

Hsieh-Yee, I., and Coogan, J. (2007, March/April). *A Comparative Analysis of Google Scholar and Academic Search Premier.* Poster presented at the Association of College & Research Libraries 13th National Conference, Baltimore, Maryland.

Vellucci, S. L. and Hsieh-Yee, I. (2007, March/April) *They Didn't Teach Me That in Library School! Building a Digital Teaching Commons to Enhance Metadata Teaching, Learning and Research*. On-site presentation and Webcast by Elluminate. A contributed paper presented at the Association of College & Research Libraries 13th National Conference, Baltimore, Maryland. The acceptance rate for contributed paper was 20%. This paper was one of 10 conference papers chosen for live webcast during the conference.

Moen, W., Hsieh-Yee, I. and Vellucci, S.L. (2007, January) *A DSpace Foundation for a Teaching & Research Commons: The Metadata Education and Research Information Commons.* A poster session presented at the Open Repositories Conference 2007, San Antonio, Texas.

Tang, R., Hsieh-Yee, I., and Zhang, S. (2006, November) *User Perception of MetaLib Combined Search*. Paper presented at the Annual Meeting of the American Society for Information Science and Technology, Austin, Texas, Nov. 2006.

Hsieh-Yee, I. (2006, November). *Federated Searching: User Perceptions, System Design, and Library Instructions.* Paper presented at the Annual Meeting of the American Society for Information Science and Technology, Austin, Texas. (Panel organizer, moderator, presenter).

Hsieh-Yee, I. (2006, November). *Building a Digital Teaching Commons to Enhance Teaching and Learning: The MERIC Experience and Challenges.* Paper presented at the Annual Meeting of the American Society for Information Science and Technology, Austin, Texas. (Panel organizer, moderator, presenter)

Hsieh-Yee, I. (2006, September). *Search Performance of Google Scholar and Academic Search Premier.* Paper presented at the ERIC Publishers Meeting, Washington, D.C.

Hsieh-Yee, I., Zhang, S., and Rong Tang, R. (2006, June). *User Perceptions of a Federated Search System*. Poster presented at Joint Conference on Digital Libraries, Chapel Hill, North Carolina.

Hsieh-Yee, I. and Zhang, S. (2006, June). *Preparing Users for Federated Search: Implications of a MetaLib User Perceptions Study.* Paper presented at the 2006 Ex Libris User Groups of North America Conference, Knoxville, Tennessee.

Hsieh-Yee, I. (2006, January). *MERIC Organizations and Navigation*. Paper presented at the 2006 ALISE Annual Conference, San Antonio, Texas.

Hsieh-Yee, I. (2006, January). *Metadata and Cataloging Education: Recommended Competencies.* Paper presented at the 2006 ALISE Annual Conference, San Antonio, Texas.

Hsieh-Yee, I. (2005, November). *Digital Library Evaluation: Progress & Next Steps.* Presentation at the Annual Meeting of the American Society for Information Science & Technology, Charlotte, North Carolina.

Hsieh-Yee, I. (2005, August). *Providing Access to Digital Content: Issues for DL Managers.* Presentation at MDK12 Digital Library Steering Committee Meeting, Columbia, Maryland.

Hsieh-Yee, I. (2005, April). *Enhancing Teaching and Learning: The Role of School Library Media Specialists.* Presentation at Meeting of the Baltimore County Public School System School Media Specialists, Baltimore, Maryland.

Hsieh-Yee, I. (2005, January). *Subject Access and Users: Insights & Inspirations from Marcia J. Bates.* Paper presented at the Historical Perspectives SIG, 2005 Conference of the Association for Library and Information Science Education, Boston, Massachusetts.

Hsieh-Yee, I. (2005, January). *Electronic Resource Management: Practice, Employer Expectations, & CE Interests.* Paper presented at Technical Services Education SIG, 2005 Conference of the Association for Library and Information Science Education, Boston, Massachusetts.

Hsieh-Yee, I. (2004, October). *Library Professionals for the Digital Age: Competencies & Preparation.* Paper presented at Bibliographic Access Management Team meeting, Library of Congress, Washington, D.C.

Hsieh-Yee, I. (2004, January). *Cataloging and metadata expertise for the digital era.* Presented at Preparing 21st Century Cataloging and Metadata Professionals: A Workshop for Educators and Trainers, San Diego and sponsored by ALCTS, ALISE, LC, and OCLC.

Hsieh-Yee, I. (2004, January). *Educating catalogers for the digital era.* Paper presented at the Technical Services SIG, 2004 Conference of the Association for Library and Information Science Education, San Diego.

Hsieh-Yee, I. (2003, July). *Cataloging Education for the 21st Century.* A presentation at the Library of Congress, Washington, D.C.

Hsieh-Yee, I. (2002, January) *Metadata Education and Research Priorities: A Delphi Study of Metadata Experts.* Presentation at the 2002 Conference of the Association for Library and Information Science Education, New Orleans.

Hsieh-Yee, I. (2001, November). *A Delphi Study of Metadata: Preliminary Findings.* Poster session at the 2001 Annual Meeting of the American Society for Information Science & Technology, Washington, D.C.

Hsieh-Yee, I. (2001, June). *Resources on Asian American Children: Analysis of Retrieval by Search Engines and WorldCat.* Presentation at the National Conference on Asian Pacific American Librarians, San Francisco.

Hsieh-Yee, I. (2001, January). *Delphi Study on Metadata: Project Design.* Presentation at Research Awards Session, Association for Library & Information Science Education, Washington, D.C.

Hsieh-Yee, I. (2000, May). *Web Search Behavior Research: Progress and Implications.* Presentation at the Symposium on Evaluating Library and Information Science Research, University of Wisconsin-Madison, Madison, Wisconsin.

Hsieh-Yee, I. (2000, March). *ERIC User Services: Evaluation in a Decentralized Environment.* Presentation at the National ERIC Joint Directors/Technical Meeting, Arlington, Virginia.

Hsieh-Yee, I. (2000, January). *Enhancing Learning with Web Technology.* Presentation at Faculty Conversations, Catholic University of America, Washington, D.C.

Hsieh-Yee, I. (2000, January). *From Surrogates to Objects: CUA's Approaches to Organizing Electronic Resources.* Paper presentation at the Annual Conference of the Association for Library and Information Science Education, San Antonio, Texas.

Yee, P., and Hsieh-Yee, I. (1997, November). *Individual Differences in Search Behavior on the WWW.* A poster session presented at the 38th Annual Meeting of the Psychonomic Society, Philadelphia, Pennsylvania.

Hsieh-Yee, I. (1997, April). *Research + Marketing + Preparation = Job!* Presented at the "Workshop on Resume and Interview Techniques," Special Libraries Association, Student Chapter, Catholic University of America, Washington, D.C.

Hsieh-Yee, I. (1997, February). *Creating CyberCatalogers: Education and Training.* Presentation at ALA's Midwinter Meeting, Washington, D.C.

Hsieh-Yee, I. (1997, February). *Search Tactics of Web Users in Searching for Texts, Graphics, Known Items and Subjects: A Search Simulation Study.* Presented at the Conference of the Association for Library and Information Science Education, Washington, D.C.

Hsieh-Yee, Ingrid. "Beginning Your Special Library/Information Center Career." Presented at SLA's "Career Day," Jan. 11, 1997, Catholic University of America.

Hsieh-Yee, I. (1996, September). *The Roles of Library and Information Scientists in Managing Electronic Information.* Presentation at Hamilton College, Clinton, New York.

Hsieh-Yee, I. (1996, May). *The Future of Cataloging as a Profession.* Presented at "The Cataloging Forum, Library of Congress, Washington, D.C.

Hsieh-Yee, I. (1994, October). *The Impact of the Internet on OPACs.* Presented at the Third Workshop on User Interfaces for OPACs, Library of Congress, Washington, D.C.

Reports

Hsieh-Yee, I., with Knowledge Management Competencies and Performance Action Group of the Federal Knowledge Management Initiative. "From Knowledge Management Competencies to Improved Organizational Performance." April 9, 2009.

Hsieh-Yee, I., with Knowledge Practices Action Group of the Federal Knowledge Management Initiative. "KM Practice in Government Agencies: Findings and Recommendations." April 9, 2009.

Hsieh-Yee, I. "Delphi Study on Metadata." 2001. Three quarterly reports submitted to the Association for Library and Information Science Education.

Hsieh-Yee, I. "College Students' Information Channels: Patterns of Use and Possible Factors in Channel Selection." 1995. Submitted to the Catholic University of America.

Hsieh-Yee, I. "The Information-Seeking Patterns of Scholars and Their Use of an Online Information System." 1994. Submitted to the Council on Library Resources.

Book Reviews

Review of *The Measurement and Evaluation of Library Services*, by Sharon L. Baker and F. Wilfrid Lancaster. *Information Processing and Management* 30 (1994): 450-52.

Review of *Subject Access to Films and Videos, by Sheila S. Intner and William E. Studwell; and Cataloging Unpublished Nonprint Materials*, by Verna Urbanski with Bao Chu Chang and Bernard L. Karon. *Information Processing and Management* 30 (1994): 449-50.

Review of *Automated Information Retrieval in Libraries: A Management Handbook*, by Vicki Anders. *Journal of Library and Information Science* 19 (1993): 98-100.

Review of *Full Text Databases*, by Carol Tenopir and Jung Soon Ro. *Information Processing and Management* 28 (1992): 667-68.

Review of *Descriptive Cataloging for the AACR2R And USMARC: A How-to-Do It Workbook*, by Larry Millsap and Terry Ellen Ferl. *Information Processing and Management* 28 (1992): 809-11.

Review of *MARC Manual: Understanding and Using MARC Records*, by Deborah J. Byrne. *Information Processing and Management* 28 (1992): 537-38.

Service

Professional Associations and Societies

- Library of Congress. RDA Training Program for the Profession. Co-authored with Tim Carlton. 2013-2014.
- 2014 Digital Preservation Outreach & Education Survey. Contributed to the design of the survey, 2014.
- National Digital Stewardship Alliance. Outreach Committee. 2011-2014.
- National Digital Stewardship Residency Program. Advisory Group, 2012-2013.
- FEDLINK Health Information Technology Advisory Council, 2011-2015.

- 2012 Joint Conference on Digital Libraries. Program Planning Committee, Pre-Conference Proposals Review Committee, 2012
- Catholic Research Resources Alliance. Five-Year Strategic Plan Task Force, 2011-2012
- Institute of Museum and Library Services. Grant reviewer. 2004, 2005, 2010.

- Association for Library and Information Science Education.
  * ALISE Bodan Wynar Research Paper Award Committee, 2015, 2016, 2017
  * ALISE Eugene Garfield Dissertation Award Competition, Jury, 2013, 2014
  * ALISE Research Grant Competition Committee. Chair, 2012
  * Pratt-Severn Faculty Innovation Award. Chair, 2009, 2010
  * ALISE Doctoral Poster Jury, 2012
  * "Information Organization Competencies for the 21st Century" Discussion session leader. 2009 Conference of the Association for Library and Information Science Education.
  * Assisted Technical Services SIG Convener in organizing a program, ""Building a Metadata Education and Research Community through MERIC (Metadata Education and Research Information Commons): Demo and Stakeholder Input" for the 2008 ALISE conference.
  * Association for Library Collections and Technical Services/Association for Library and Information Science Education (ALCTS/ALISE) Metadata Education and Research Information Center (MERIC) Advisory Board, Co-Chair (with Sherry Vellucci), 2005-2007. Chair, 2008-2009 (leading the effort to build MERIC, a repository and collaborative space for metadata educators, practitioners, and researchers)
  * Technical Services SIG, Convener, 2004-2005. Organized a program on "Electronic Resources Management: Current Practices, Employer Expectations, and Teaching Strategies" for the 2005 conference in Boston, Massachusetts.
  * Technical Services SIG, Convener, 2003-2004. Organized a program on "Organizing Information with Metadata: Desired Competencies and Teaching Innovations" for the 2004 conference.
  * Technical Services SIG, Convener, 1999-2000. Organized a program on "Teaching the Organization of Electronic Resources" for the 2000 conference.
  * Curriculum SIG, Co-convener (with Sibyl Moses), 1996-97. Organized a program on "Government Information Policy" for the 1997 conference.

- American Society for Information Science & Technology.
  * Reviewer, Conference program panel submissions and poster submissions, 2005, 2006, 2007, 2009, 2011, 2012, 2013, 2014, 2015, 2016, 2017
  * Nomination Committee, 2009-2011
  * Information Science Education Special Interest Group. American Society for Information Science and Technology. Chair-Elect, 2007-2008. Chair 2008-2009.
  * Committee on Information Science Education. 1999-2006.
  * Committee on Information Science Education. Organizing Committee for an orientation program for students at ASIS annual meetings, 1999-2001
  * Committee on Information Science Education. Sub-committee on Student Welfare (focusing on issues related to master's education), 1998-2001

* SIG ED. Organizing Committee for the "Seminar on Research and Career Development" for junior researchers. 1995-96 (chair), 1997-2001
* ISI Doctoral Dissertation Proposal Scholarship Jury, 1997; 2001, 2002
* Pratt-Severn Best Student Research Paper Award Jury. Chair. 1997
* 1998 Midyear Meeting (referee of contributed papers), 1997
* Organizer and moderator of the ASIS Doctoral Forum and the Doctoral Research Seminar 1994-1995
* SIG Human Computer Interaction. Chair-Elect, Chair, 1993-1995
* Doctoral Forum Award Jury, 1995
* Best Student Paper Award Jury, 1995

- American Library Association.
  * Committee on Accreditation, External Review Panelist, 2009- (site visiting team 2013-2014; site visiting team 2016-2017)
  * Association for Library Collections and Technical Services Task Force on Competencies and Education for a Career in Cataloging, member, 2008-2009
  * Facilitator for "What They Don't Teach in Library School: Competencies, Education and Employer Expectations for a Career in Cataloging," an Association for Library Collections and Technical Services Preconference, June 22, 2007 in Washington, D.C. Also a local liaison for bringing this program to the Catholic University of America.
  * Facilitator for a discussion on "Effect of Electronic Resources on Technical Services" at ALA's Midwinter Meeting held in Feb. 1997 in Washington, D.C.
  * International Relations Committee, Subcommittee Task Force for IFLA and China, 1994-1997

- Virginia Association of School Librarians. Scholarships and Awards Committee. 2010-2012

- Federal Knowledge Management Initiative, Knowledge Management Practices Action Group. Member. 2009 (leading the effort to build a knowledge management repository)
- Federal Knowledge Management Initiative, Knowledge Management Competencies & Learning Action Group. Member. 2009 (developing an action plan for helping government knowledge workers and government agencies to develop knowledge management competencies)
- National Center for Education Statistics. Technical Review Panel. 2008.
- External evaluator for a case of promotion to full professorship. University of Tennessee. 2008.
- National Information Standards Organization (NISO). Advisory Board, Revision of "IMLS Framework of Guidance for Building Good Digital Collections," 2004, 2007.
- Library of Congress, Bibliographic Control of Web Resources: A Library of Congress Action Plan. Principal Investigator of Action Item 5.1, focusing on cataloging and metadata education for students and new librarians, 2002-2003. (worked with the Association for Library Collections and Technical Services, Education Task Force)

- Chinese American Librarians Association
  * Chinese American Librarians Association Outstanding Library Leadership Award in Memory of Dr. Margaret Chang Fung, Award Committee, 2016-2017
  * Achievement Award Jury, 2000-2001
  * CALA Goal 2000 Task Force, 1997
  * Scholarship Committee, 1995, 1996-1997 (chair)
  * Board of Directors, 1994-1997
  * Publication Committee, 1993-1995
  * International Relations Committee, 1993-1996

- Sailor[SM] Assessment Advisory Group (An impact study of Sailor, Maryland's Public Information Network), 1995

- Editorial boards
  Journal of Library & Information Science. Editorial Board, 2012-
  Chinese American Librarians Association, *Occasional Papers Series*. Editorial Board, 2009-2016.
  Library Quarterly. Editorial Board, 2003-2008
  Bulletin of the Medical Library Association, 1994-97
  Newsletter editor for the Chinese American Librarians Association, 1989-92

- Referee for the following journals
  Information Processing and Management
  Journal of Digital Information
  Journal of Education for Library and Information Science
  Journal of Information Science
  Journal of Library & Information Science
  Journal of Library Metadata
  Journal of the American Society for Information Science & Technology
  Library and Information Science Research
  Library Quarterly

- Expert reviewer, "Digital Library" course, Evaluation module, University of North Carolina, Chapel Hill, 2007-2008.
- Expert reviewer, "Information Organization" course, University of Michigan, Ann Arbor. 2007.

Catholic University of America

- Academic Senate Committee on Appointments and Promotions, 2005-2008
- Academic Senate Committee on Committees and Rules, 2009-2012, 2019-2022
- Academic Senate representative (for School of Arts & Sciences), 2017-2020

- Academic Senate representative (for School of Library & Information Science) 2003-2012
- Academic Senate Library Committee, Interim Chair (2007), Member, 2008-2012
- Academic Senate Committee on Computing, 1995-2003

- President's Administrative Council, 2010-2012
- Deans' Council, 2010-2012
- Academic Leadership Group, 2010-2012
- Graduate Board, 2010-2012

- School of Arts & Sciences, Committee on Appointments and Promotions, 2015-2019
- School of Arts & Sciences, Academic Council, 2015-2016.
- School of Arts & Sciences, Ordinary Professor Group, 2013-
- Doctoral Dissertation Defense Committee, Chair, Dept. of Psychology, 2016, 2017, 2018, 2019, 2020, 2021, 2022
- Doctoral Dissertation Defense Committee, Chair, Dept. of Education, 2014, 2015, 2017, 2018, 2019
- Doctoral Dissertation Defense Committee, Chair, National Catholic School of Social Services, 2019, 2020, 2021
- Doctoral Dissertation Defense Committee, Chair, School of Nursing, 2006, 2008
- CUA Scholarly Communication Project Team, Member (2007), Chair, 2008-2009
- Dean Search Committee, 1992-1994, 1998-1999, 2002-2003, 2006-2007
- Fulbright Review Panel, 2006
- CUA Service Learning Advisory Board, 2001-2002
- CUA Faculty Conversations on Enhancing Teaching and Learning through Technology, Planning Group, 1999-2001
- CUA Initiative on Technology and Teaching, 1998-2001

Dept. of Library and Information Science

- Admissions Committee, 2007-2009, Chair 2010-2012, Member 2013-2015, Member 2018-2022
- Appointments and Promotions Committee, 1991-2022
- Student advisement, 1990-2022
- 4+1 Bachelor to Master program, Program Advisor, 2014-2022
- LIS Minor, Advisor, 2017-2022
- Curriculum Committee, 2020-21 (Member), 2021-2022 (Chair)
- Scholarship & Awards Committee, 2016-2018, 2020-21
- Comprehensive examination editor, 2020-21
- Symposium, Colloquium, Lecture Series Committee, May 2018-May 2020 (Chair)
- Symposium and Colloquium Committee, fall 2016-May 2018
- Bridging the Spectrum Symposium session moderator, 2018, 2019, 2020, 2021

- Organization of Information Core Course Outcomes Assessment, 2019, 2020
- The Information Professions in Society Core Course Outcomes Assessment, 2019, 2020
- Community Services Librarianship, Course of Study Review and Revision, 2020
- Organization of Information, Course of Study Review and Revision, 2019
- Organization of Information, Competency and Course Offerings Review, 2018-2019
- Accreditation presentation, Chair, June 2015-August 2016
- Interim Co-Chair, June 2015-August 2016.
- Blended/OWL Learning Committee, spring 2016-2018
- Technology Committee, fall 2016-2017
- Comprehensive examination editor, 2016-2017, reader (every year since 1990)
- LIS Advisory Board, 2015-2016 (chair); fall 2016- May 2018 (member)
- Committee on Planning and Assessment, 2015-2016 (chair)
- Senior Faculty Committee, 2014-2016.
- Accreditation Steering Committee, 2014-2016 (Chair, 2015-2016)
- Accreditation Students Standard Committee, co-chair, 2014-2016
- Accreditation Mission, Goals, and Objectives Standard Committee, co-chair, 2014-2016
- Accreditation Curriculum Standard, member 2014-2-16
- Accreditation Administration and Finance Standard, member 2014-2016
- Cultural Heritage Information Management Project (IMLS-funded), Co-PI, 2012-2015
- Cultural Heritage Information Management Forum (scheduled for June 2015), Co-Organizer, 2013-2015
- Health Information Technology Interim Review Committee, 2015 (chair)
- Health Sciences Librarianship Advisory Group, 2015- (chair)
- Comprehensive examination editors, 2013-2014, 2016-2017
- National Digital Stewardship Alliance liaison, 2011-2014
- Advisory Board, Chair 2010-2012
- Academic Honesty Committee, Chair, 2008-2012
- Blended Learning Committee, 2010-2012
- Colloquium Committee, 2010-2012
- Comprehensive Examination Administration, 2010-2012
- Cultural Heritage Information Management Advisory Committee, 2010-2012 (chair), 2013-
- Curriculum Committee, 1991-2003, 2007-2009, Chair 2010-2012, member 2013-
- Curriculum Subcommittee on Comprehensive Examination, Chair 2009-2012
- Health Information Technology Advisory Board, Chair 2010-2012. Member 2013-
- Health Sciences Advisory Committee, 2009, Chair 2010-2012. Member 2013-
- HIT Expert Forum, Chair 2012. Member 2013-2015
- Health Information Technology Student Group Advisor, 2011-2012
- State Council for Higher Education of Virginia, SLIS Representative, 2010-2012
- Symposium Planning Committee, 2010-2012
- Website Management Team, Chair, 2010-2012
- Urban School Librarianship Project (IMLS-Funded), PI, 2007-2011 (chair, 2010-11)

- Failing Grades Committee, 1995-1997 (chair), 2000-2001 (chair), 2004-2005 (chair), 2007 (chair)-2011
- Faculty Search Committee, 1994-1998, 2002-2004, 2006 (chair), Fall 2007-2009, Chair fall 2009-2012
- Recruitment Committee, Chair 2010-2012
- Strategic Planning Committee, Chair 2010-2012
- Technology Committee, 2010-2012
- Accreditation Advisory Committee, 2007-2009
- Accreditation Coordinating Committee, 2007-2009
- Accreditation Steering Committee, 2007-2009
- SLIS Advisory Group, 2007-2009
- Accreditation Curriculum Standard Committee, Co-chair, 2007-2009
- Accreditation Faculty Standard Committee, Co-chair, 2007-2009
- LSC 551 Information Organization Review Team, Co-chair, 2008-2009, 2015-2016.
- Curriculum Subcommittee on Portfolios, 2009
- LSC 555 Information Systems in Libraries and Information Centers Review Team, contributor, 2008-2009
- Redesign of LSC 730 Use and Users of Libraries and Information. 2009-
- Development of a metadata institute that was taught as LSC 715 Organization of Internet Resources in 2008. The institute is being revised and will be offered in 2010 under a new course title.
- Development of lesson plans, assignments, and evaluation rubrics for LSC 606, Cataloging and Classification, for the School's NCATE accreditation. 2008
- Howard and Mathilde Rovelstad Scholarship Committee, Chair, 2004-2007
- Assistant Dean Search Committee, Chair, Fall 2007
- Liaison to the Association for Library Collections and Technical Services to bring its preconference program, Cataloging Education and Employer Expectations, to CUA during the 2007 American Library Association Annual Meeting in Washington, D.C.
- Organizer of the colloquium presentation and reception for Tamar Sadeh of Ex Libris on PRIMO June 2007
- Practicum review and design (work with potential supervisors, such as the American Indian Museum internship description revision) 2006-
- Comprehensive examinations (edits, proctoring, and grading), 1990-
- SLIS Web site redesign: Comments and suggestions. Fall 2007
- Conducted surveys of current students and alumni in preparation for the 2005 re-accreditation, 2004-2005
- Technology Committee, 1992-1999 (chair, 1996-1998), 2002-2003 (member)
- Colloquia Committee 1997-1999, 2002-2003.
- Advisor of the CUA Student Chapter of the American Society for Information Science and Technology, 2002-2003
- Visiting Professor Search Committee, 1999, 2000, 2001
- Leader, Participation in the CORC experiment, 1999-2000
- Advisor of the Special Libraries Association Student Chapter, 1993-1999; the group was recognized for outstanding leadership by SLA in 1999.

- COA planning Committee, Task Force on Electronic Presentation of SLIS Reports (team leader) 1997-1998
- COA Planning Committee, Subcommittee on Technology 1996-1998
- NLM practicum coordinator, 1997-1998
- Computer Literacy Workshops: Assisted with the development and evaluation of the workshops, 1996-1998
- Leader, Participation in the InterCat project, 1995-1997

Appendix B

MARC 21 Reference Materials

Part VII: A Summary of Commonly Used MARC 21 Fields
Part VIII: A List of Other Fields Often Seen in MARC Records
Part IX: The Leader
Part X: Field 008 for Books

Part VII:

A Summary of Commonly Used MARC 21 Fields

This is a summary of the MARC 21 tags used most frequently by libraries in entering their own bibliographic records. For full listings of all MARC 21 tags, indicators, and subfield codes, see *MARC 21 Format for Bibliographic Data*.

In the explanations on these pages:

Tags -- The tags (3-digit numbers) are followed by the names of the fields they represent. In this summary, and in the *MARC 21 Format for Bibliographic Data*, if a tag can appear more than once in one bibliographic record, it is labeled repeatable (R). If it can only be used once, it is labeled non-repeatable (NR). For example, a catalog record can have several subjects, so the tags for subject added entries (6XX) are labeled repeatable (R).

Indicators -- The use of indicators is explained in fields where they are used. Indicators are one-digit numbers. Beginning with the 010 field, in every field -- following the tag -- are two character positions, one for Indicator 1 and one for Indicator 2. The indicators are not actually defined in all fields, however. And it is possible that a 2nd indicator will be used, while the 1st indicator remains undefined (or vice versa). When an indicator is undefined, the character position will be represented by the character # (for blank space).

Subfield codes -- All the data in each field (beginning with the 010 field) is divided into subfields, each of which is preceded by a delimiter-subfield code combination. The most common subfield codes used with each tag are shown. Each subfield code is preceded by the character $, signifying a delimiter. The name of the subfield follows the code.

In general, every field MUST have a subfield 'a' (**$a**). One exception that is often seen is in Field 020 (ISBN), when the ISBN information (subfield **$a**) is unavailable but the price (subfield **$c**) is known. Some subfields are repeatable. In this summary, repeatability is noted for only the more common repeatable subfields.

Examples: Examples follow the explanation for each field. For clarity, one space has been placed between the tag and the first indicator, one space has been placed between the second indicator and the first delimiter- subfield code, and one space has been inserted between the delimiter-subfield code and the subfield data.

010 Library of Congress Control Number -- (LCCN)
(NR, or Not Repeatable
Indicators undefined.

Subfield used most often:

$a -- Library of Congress control number

```
Example:    010 ##  $a ###86000988#
```

020 International Standard Book Number -- (ISBN)
(R, or Repeatable)

Indicators undefined.

Subfields used most often:

$a -- International Standard Book Number
$c -- Terms of availability (often a price)
$z -- Cancelled/invalid ISBN (R)

```
Example:    020 ##  $a 0877547637
```

040 Cataloging source -- (NR)

Indicators undefined.

Subfields used most often:

$a -- Original cataloging agency
$c -- Transcribing agency
$d -- Modifying agency (R)

```
Example:    040 ## $a DLC
                   $c DLC
                   $d gwhs
```

100 Main entry -- Personal name -- (primary author)
(NR; there can be only one main entry)

Indicator 1: Type of personal name entry element

0 -- Forename
1 -- Surname (this is the most common form)
3 -- Family name

Indicator 2 undefined.

Indicator 2 became obsolete in 1990. Older records may display 0 or 1

Subfields used most often:

$a -- Personal name
$b -- Numeration
$c -- Titles and other words associated with a name (R)
$q -- Fuller form of name
$d -- Dates associated with a name (generally, year of birth)

```
Example:    100 1# $a Gregory, Ruth W.
```

```
                          $q (Ruth Wilhelme),
                          $d 1910-
```

130 Main entry -- Uniform title -- (NR)

Indicator 1: Nonfiling characters

0-9 -- Number of nonfiling characters present (for initial articles, including spaces)

Indicator 2 undefined.

Indicator 2 became obsolete in 1990. (See 100 above.)

Subfields used most often:

$a -- Uniform title
$p -- Name of part/section of a work (R)
$l -- Language of a work
$s -- Version
$f -- Date of a work

```
Example:    130 0# $a Bible.
                   $p O.T.
                   $p Psalms.
```

240 Uniform title (NR)

Indicator 1: Uniform title printed or displayed

0 -- Not printed or displayed
1 -- Printed or displayed (most common)

Indicator 2: Nonfiling characters

0-9 -- Number of nonfiling characters present (for initial articles, including spaces)

Subfields used most often:

$a -- Uniform title
$l -- Language of a work
$f -- Date of a work

```
Example:    240 10 $a Ile mystérieuse.
                   $l English.
                   $f 1978
```

245 Title Statement (NR)

Indicator 1: Title added entry

(Should the title be indexed as a title added entry?)

0 -- No title added entry

(indicates a title main entry; i.e. no author is given)

1 -- Title added entry

(the proper indicator when an author given in 1XX; the most common situation)

Indicator 2: Nonfiling characters

0-9 -- Number of nonfiling characters present, including spaces; usually set at zero, except when the title begins with an article; e.g., for *The robe*, the second indicator would be set to 4. The letters *T*, *h*, *e*, and the space following them are then ignored in alphabetizing titles. The record will be automatically filed under "*r*" -- for *Robe*.

Subfields used most often:

$a -- Title proper
$h -- Medium (often used for non-book media)
$p -- Name of part/section of a work (R)
$b -- Reminder of title (subtitles, etc.)
$c -- Remainder of title page transcription/Statement of responsibility

```
Example:    245 14 $a The DNA story :
                   $b a documentary history of gene
                      cloning /
                   $c James D. Watson, John Tooze.
```

246 Varying form of title (R)

Indicator 1: Note/title added entry controller
1 -- Note, title added entry
3 -- No note, title added entry

Indicator 2: Type of title
-- No information provided
0 -- Portion of title
1 -- Parallel title
4 -- Cover title
8 -- Spine title

Subfield used most often:
$a -- Title proper

```
Example:    246 3# $a Four corners power review
```

250 Edition statement (NR)

Indicators undefined.

Subfield used most often:
$a -- Edition statement

```
Example:    250 ## $a 6th ed.
```

260 Publication, distribution, etc. (Imprint) (R)

Indicator 1: Sequence of publishing statements
-- No information provided

Indicator 2: Undefined

Google Exhibit 1020
Page 00149

Subfields used most often:

$a -- Place of publication, distribution, etc. (R)
$b -- Name of publisher, distributor, etc. (R)
$c -- Date of publication, distribution, etc. (R)

```
Example:    260 ## $a New York :
                   $b Chelsea House,
                   $c 1986.
```

300 Physical description (R)

Indicators undefined.
Subfields used most often:

$a -- Extent (number of pages) (R)
$b -- Other physical details (usually illustration information)
$c -- Dimensions (cm.) (R)
$e -- Accompanying material (for example, "teacher's guide" or "manual")

```
Example:    300 ## $a 139 p. :
                   $b ill. ;
                   $c 24 cm.
```

440 Series statement / Added entry--Title

This field was made obsolete in 2008 to simplify the series statement. See 490 and 830.

490 Series statement (No added entry is traced from field) (R)

Indicator 1: Specifies whether series is traced (whether an 8XX tag is also present)

0 -- Series not traced
1 -- Series traced (8XX is in record)

Indicator 2 undefined.
Subfield used most often:

$a -- Series statement (R)
$v -- Volume number (R)

```
Example:    490 1# $a Colonial American craftsmen
```

500 General note (R)

Indicators undefined.
Subfield used most often:

$a -- General note (Used when no specialized note field has been defined for the information. Examples: Notes regarding the index; the source of the title; variations in title; descriptions of the nature, form, or scope of the item.)

Google Exhibit 1020
Page 00150

```
Example:    500 ##  $a Includes index.
```

504 Bibliography, etc. note (R)

Indicators undefined.

Subfield used most often:

$a -- Bibliography, etc. note

```
Example:    504 ## $a Includes bibliographical
                      references.
```

505 Formatted contents note (R)

Indicator 1: Type of contents note

0 -- Complete contents

1 -- Incomplete contents (used with multivolume set when some volumes are not yet published)

2 -- Partial contents

Indicator 2: Level of content designation

-- Basic

Subfield used most often:

$a -- Formatted contents note

```
Example:    505 0# $a Pride and prejudice -- Emma
                      -- Northanger Abbey.
```

520 Summary, etc. note (R)

Indicator 1: Display constant controller

-- Summary

1 -- Review

2 -- Scope and content

3 -- Abstract

Indicator 2 undefined

Subfields used most often

$a -- Summary, abstract, or annotation

$b -- Expansion of summary note

```
Example:    520 ## $a This basic guide to parliamentary
                      procedure tells how to conduct
                      and participate in a meeting
                      properly.
```

600 Subject added entry -- Personal name (R)

Indicator 1: Type of personal name entry element

0 -- Forename

1 -- Surname (this is the most common form)
3 -- Family name

Indicator 2: Subject heading system/thesaurus (identifies the specific list or file which was used)

0 -- Library of Congress Subject Headings
1 -- LC subject headings for children's literature
2 -- Medical Subject Headings
3 -- National Agricultural Library subject authority file
4 -- Source not specified
5 -- Canadian Subject Headings
6 -- Répertoire de vedettes-matière
7 -- Source specified in subfield $2

(Note regarding Sears subject headings: The MARC 21 format does not provide an assigned indicator for Sears subject headings. Therefore, an indicator of 7 is used, and the MARC defined code "sears" is placed in subfield $2.)

Subfields used most often:

$a -- Personal name (surname and forename)
$b -- Numeration
$c -- Titles and other words associated with a name (R)
$q -- Fuller form of name
$d -- Dates associated with a name (generally, year of birth)
$t -- Title of a work
$v -- Form subdivision (R)
$x -- General subdivision (R)
$y -- Chronological subdivision (R)
$z -- Geographic subdivision (R)
$2 -- Source of heading or term (used with 2nd indicator of 7)

```
Example:    600 10 $a Shakespeare, William,
                   $d 1564-1616
                   $x Comedies
                   $x Stage history.
```

```
Example:    600 10 $a Shakespeare, William,
                   $d 1564-1616
                   $x Knowledge
                   $z Rome
                   $v Congresses.
```

Notice that subfields $v, $x, and $z in the 600 field are repeatable. Subfields $v, $x, $y, and $z do not have to be in alphabetical order. They will be in the order prescribed by the instructions given by the subject heading system.

610 Subject added entry -- Corporate name (R)

*Indicator 1:*Type of corporate name entry element

0 -- Inverted name (not used with AACR2)
1 -- Jurisdiction name
2 -- Name in direct order

Indicator 2: Subject heading system/thesaurus.
See indicator 2 under 600

Subfields used most often:
$a -- Corporate name or jurisdiction name as entry element
$b -- Subordinate unit (R)
$v -- Form subdivision (R)
$x -- General subdivision (R)
$y -- Chronological subdivision (R)
$z -- Geographic subdivision (R)
$2 -- Source of heading or term (used with 2nd indicator of 7)

```
Example:    610 10 $a United States.
                   $b Army Air Forces
                   $v Biography.
```

650 Subject added entry -- Topical term (Most subject headings fit here.) (R)

Indicator 1: Level of subject
-- No information provided

Indicator 2: Subject heading system/thesaurus
(identifies the specific list or file which was used)
0 -- Library of Congress Subject Headings
1 -- LC subject headings for children's literature
2 -- Medical Subject Headings
3 -- National Agricultural Library subject authority file
4 -- Source not specified
5 -- Canadian Subject Headings
6 -- Répertoire de vedettes-matière
7 -- Source specified in subfield $2

Note regarding Sears subject headings: The MARC 21 format does not provide an assigned indicator for Sears subject headings. Therefore, an indicator of 7 is used, and the MARC defined code "sears" is placed in subfield $2.)

Subfields used most often:
$a -- Topical term
$v -- Form subdivision (R)
$x -- General subdivision (R)
$y -- Chronological subdivision (R)
$z -- Geographic subdivision (R)
$2 -- Source of heading or term used with 2nd indicator of 7)

```
Example:    650 #0 $a Theater
                   $z United States
                   $v Biography
                   $v Dictionaries.
```

Notice that subfields $v, $x, and $z in the 650 field are repeatable. Subfields $v, $x, $y, and $z do not have to be in alphabetical order. They will be in the order prescribed by the instructions given by the subject heading system.

651 Subject added entry -- Geographic name (R)

Indicator 1: undefined.
Indicator 2: Subject heading system/thesaurus.
See indicator 2 under 600
Subfields used most often:
$a -- Geographic name
$v -- Form subdivision (R)
$x -- General subdivision (R)
$y -- Chronological subdivision (R)
$z -- Geographic subdivision (R)
$2 -- Source of heading or term (used with 2nd indicator of 7)

```
Example:    651 #0 $a United States
                   $x History
                   $v Chronology.
```

Notice that subfields $v, $x, and $z in the 651 field are repeatable. Subfields $v, $x, $y, and $z do not have to be in alphabetical order. They will be in the order prescribed by the instructions given by the subject heading system.

700 Added entry -- Personal name (R)

Indicator 1: Type of personal name entry element
0 -- Forename
1 -- Surname (this is the most common form)
3 -- Family name
Indicator 2: Type of added entry
-- No information provided (most common; co-authors, editors, etc.)
2 -- Analytical entry (The values for Indicator 2 changed in 1994 with Format Integration, and older records may display additional values. An analytical entry involves an author/title of an item contained in a work.)
Subfields used most often:
$a -- Personal name
$b -- Numeration
$c -- Titles and other words associated with a name (R)
$q -- Fuller form of name
$d -- Dates associated with a name (generally, year of birth)
$e -- Relator term (such as ill.) (R)
$4 -- Relator code (R)

```
Example:    700 1# $a Baldridge, Letitia.
```

710 Added entry -- Corporate name (R)

Indicator 1: Type of corporate name entry element
- 0 -- Inverted name (not used with AACR2)
- 1 -- Jurisdiction name
- 2 -- Name in direct order

Indicator 2: Type of added entry.
- See Indicator 2 under 700
- # -- No information provided
- 2 -- Analytical entry

Subfields used most often:
- **$a** -- Corporate name or jurisdiction name as entry element
- **$b** -- Subordinate unit (R)

```
Example:    710 2# $a Sunburst Communications (Firm)
```

740 Added entry -- Uncontrolled related/analytical title (R)

Indicator 1: Nonfiling characters
- 0-9 -- Number of nonfiling characters present (for initial articles, including spaces)

*Indicator 2:*Type of added entry. See Indicator 2 under 700
- # -- No information provided
- 2 -- Analytical entry
- (This field was redefined in 1994 with Format Integration. Prior to 1994, the field was also used for variant titles, such as a different wording on a spine title. In records created since Format Integration, those variant tiles appear in a 246 field.)

Subfield used most often:
- **$a** -- Title

```
Example:    740 02 $a Uncle Vanya.
```

800 Series added entry -- Personal name (R)

Indicator 1: Type of personal name entry element
- 0 -- Forename
- 1 -- Surname
- 3 -- Family name

Indicator 2 undefined.

Subfields used most often:
- **$a** -- Personal name
- **$b** -- Numeration
- **$c** -- Titles and other words associated with a name (R)
- **$q** -- Fuller form of name

Google Exhibit 1020
Page 00155

$d -- Dates associated with a name (generally, year of birth)
$t -- Title of a work (the series)
$v -- Volume number

```
Example:    800 1# $a Fisher, Leonard Everett.
                   $t Colonial American craftsmen.
```

830 Series added entry -- Uniform title (R)

Indicator 1 undefined.
Indicator 2: Nonfiling characters
0-9 -- Number of nonfiling characters present (for initial articles, including spaces)
Subfield used most often:
$a -- Uniform title
$v -- Volume number

```
Example:    830 #0 $a Railroads of America (Macmillan)
```

[Back to Top of Page]

Part VIII:

A List of Other Fields Often Seen in MARC Records

001	Control number
003	Control number identifier
005	Date and time of latest transaction
006	Fixed-length data elements -- additional material characteristics
007	Physical description fixed field
008	Fixed length data elements (See Part X)
022	International Standard Serial Number (ISSN)
037	Source of acquisition
041	Language code
043	Geographic area code
050	Library of Congress call number
060	National Library of Medicine call number
082	Dewey Decimal classification number (the one recommended by the Library of Congress; locally-assigned call numbers may appear elsewhere)
110	Main entry -- Corporate name (less frequent under AACR2 rules)
256	Computer file characteristics
263	Projected publication date (indicates a CIP -- Cataloging in Publication -- record)

Google Exhibit 1020
Page 00156

306 Playing time
508 Creation/production credits note
510 Citation/references note (review sources)
511 Participant or performer note
521 Target audience note (first indicator: 0 = reading grade level, 1 = interest age level, 2 = interest grade level, 3 = special audience characteristics, 4 = motivation interest level)
530 Additional physical form available note
538 System details note
586 Awards note
656 Index term -- Occupation
730 Added entry -- Uniform title
852 Location
856 Electronic location and access
9XX Reserved for local use. (They are used by vendors, systems, or individual libraries to exchange additional data)

[Back to Top of Page]

Part IX:

The Leader

There are 24 positions in the Leader, numbered from 00 to 23. For fuller explanation, see the *MARC 21 Format for Bibliographic Data*.

00-04 Record length (calculated by the computer for each record)
05 Record status
- a = increase in encoding level
- c = corrected or revised
- d = deleted
- n = new
- p = increase in encoding from prepublication (previous CIP)

06 Type of record
- a = language material
- c = printed music
- d = manuscript music
- e = cartographic material
- f = manuscript cartographic material
- g = projected medium
- i = nonmusical sound recording
- j = musical sound recording
- k = 2-dimensional nonprojectable graphic
- m = computer file
- o = kit
- p = mixed materials
- r = 3-dimensional artifact or naturally occurring object
- t = manuscript language material

07 Bibliographic level
- a = monographic component part
- b = serial component part

c = collection
d = subunit
i = integrating resource
m = monograph/item
s = serial

08 Type of control
# = no specified type
a = archival

09 Character coding scheme
# = MARC-8
a = UCS/Unicode

10 Indicator count (always "2")

11 Subfield code count (always "2")

12-16 Base address of data (calculated by the computer for each record)

17 Encoding level
# = full level
1 = full level, material not examined
2 = less-than-full level, material not examined
3 = abbreviated level
4 = core level
5 = partial (preliminary) level
7 = minimal level
8 = prepublication level (CIP)
u = unknown
z = not applicable

18 Descriptive cataloging form
# = non-ISBD
a = AACR2
i = ISBD
u = unknown

19 Multipart resource record level
# = Not specified or not applicable
a = Set
b = Part with independent title
c = Part with dependent title

20 Length of the length-of-field portion (always "4")

21 Length of the starting-character-position portion (always "5")

22 Length of the implementation-defined portion (always "0")

23 Undefined (always "0")

[Back to Top of Page]

Part X:

Field 008 for Books

Field 008 is used for Fixed Length Data Elements ("Fixed Field Codes"). There are 40 character positions in field 008, numbered from 00-39. Undefined positions must contain either a blank (#) or a fill character (|). Positions 00-17 and 35-39 are defined the same way for all media.

The information shown here for positions 18-34 applies only to books. For explanation of all the positions below and for positions 18-34 for other media, see the *MARC 21 Format for Bibliographic Data*.

Note that field 008 has no indicators or subfield codes.

00-05 Date entered on file (YYMMDD),
where Y=year, M=month, and D=day

06 Type of date/publication status:
- b = no dates given; B.C. date involved
- e = detailed date
- s = single known date/probable date
- m = multiple dates
- r = reprint/reissue date (Date 1) and original date (Date 2)
- n = dates unknown
- q = questionable date
- t = publication date and copyright date
- | = no attempt to code

07-10 Date 1/beginning date of publication

11-14 Date 2/ending date of publication

Date fields contain the year(s) of publication. The type of date(s) in these elements are specified in fixed field element 06: Type of date/publication status. (For further details, see the field 008 description in the *MARC 21 Format for Bibliographic Data*.)

15-17 Place of publication, production, or execution
For example:
- pk# = Pakistan
- cau = California (US)

(For a full list of codes used in these positions, see the *MARC Code List for Countries*.)

18-21 Illustrations (up to 4 codes):
- # = no illustrations
- a = illustrations
- b = maps
- c = portraits
- d = charts
- e = plans
- f = plates
- g = music
- h = facsimiles
- i = coats of arms
- j = genealogical tables
- k = forms
- l = samples
- m = phonodisc, phonowire, etc.
- o = photographs
- p = illuminations
- | = no attempt to code

22 Target audience:
- # = unknown or not specified
- a = preschool

b = primary
c = pre-adolescent
d = adolescent
e = adult
f = specialized
g = general
j = juvenile
| = no attempt to code

23 Form of item:
# = none of the following
a = microfilm
b = microfiche
c = microopaque
d = large print
f = braille
r = regular print reproduction
s = electronic
| = no attempt to code

24-27 Nature of contents (up to 4):
# = no specified nature of contents
a = abstracts/summaries
b = bibliographies (is one or contains one)
c = catalogs
d = dictionaries
e = encyclopedias
f = handbooks
g = legal articles
i = indexes
j = patent document
k = discographies
l = legislation
m = theses
n = surveys of literature
o = reviews
p = programmed texts
q = filmographies
r = directories
s = statistics
t = technical reports
u = standards/specifications
v = legal cases and notes
w = law reports and digests
z = treaties
| = no attempt to code

28 Government publication:
# = not a government publication
i = international intergovernmental
f = federal/national
a = autonomous or semi-autonomous component
s = state, provincial, territorial, dependent, etc.
m = multistate
c = multilocal

l = local
z = other type of government publication
o = government publication -- level undetermined
u = unknown if item is government publication
| = no attempt to code

29 Conference publication:
0 = not a conference publication
1 = conference publication
| = no attempt to code

30 Festschrift:
0 = not a festschrift
1 = festschrift
| = no attempt to code

31 Index:
0 = no index
1 = index present
| = no attempt to code

32 Undefined (since 1990) (Earlier records may contain the values 0 or 1)
= Undefined
| = no attempt to code

33 Literary form:
0 = not fiction (not further specified)
1 = fiction (not further specified)
c = comic strips
d = dramas
e = essays
f = novels
h = humor, satires, etc.
i = letters
j = short stories
m = mixed forms
p = poetry
s = speeches
u = unknown
| = no attempt to code

34 Biography:
# = no biographical material
a = autobiography
b = individual biography
c = collective biography
d = contains biographical information
| = no attempt to code

35-37 Language:
A three-letter code. For example: eng fre ger spa rus ita

(For a full list of codes used in these positions, see the *MARC Code List for Languages*.)

38 Modified record:
# = not modified
x = missing characters (because of characters unavailable in MARC character set)
s = shortened
d = "dashed-on" information omitted

r = completely romanized/printed cards in script
o = completely romanized/printed cards romanized
| = no attempt to code

39 Cataloging source:
# = national bibliographic agency
c = cooperative cataloging program
d = other sources
u = unknown
| = no attempt to code

[Back to Top of Page]

[Back to Table of Contents] -- [Continue to Part 11]

Library of Congress

Library of Congress Help Desk (10/27/2009)