

QUANDO LA LINGUISTICA INCONTRA L'INFORMATICA: UNA RIFLESSIONE TERMINOLOGICA

Gabriella Pardelli*, Stefania Biagioni**

* Istituto di Linguistica Computazionale "Antonio Zampolli" – CNR, Pisa Italy

** Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo" – CNR, Pisa, Italy

Abstract

Le ricerche linguistiche, a partire dal secondo dopoguerra, hanno avuto un ritmo di evoluzione e di espansione molto rapido, grazie anche ai metodi di analisi introdotti, come l'uso dei metodi statistici o quantitativi nello studio delle lingue e delle opere letterarie. Nacquero nuovi settori di applicazione, la linguistica incontrò altre scienze e l'interdisciplinarietà venne sempre più praticata fino a diventare necessaria. L'introduzione di sistemi di automazione nelle analisi linguistiche vide la nascita della Linguistica Computazionale (LC) che mise in connessione lo studio della lingua con l'ausilio dell'elaboratore elettronico. Dalla fine degli anni '40 all'inizio degli anni '60 gli utilizzi del calcolo elettronico per l'elaborazione di dati linguistici si articolano in due filoni principali: i) gli spogli elettronici dei testi che diedero impulso alla lessicografia computazionale, avviata da Padre Roberto Busa nel 1951 con la compilazione delle concordanze di Tommaso d'Aquino; ii) i tentativi di traduzione automatica (TA), in inglese machine translation (MT), avviati da Weaver nel 1949 con la pubblicazione del memorandum "Translation". La traduzione automatica divenne da subito nucleo e centro di spinta della Linguistica Computazionale utilizzando il calcolatore per trasportare un testo da una lingua naturale all'altra. Il presente articolo sintetizza il rapido percorso della LC e la necessità di elaborare in fretta una terminologia adeguata alla neonata disciplina. Fornisce, inoltre, indicazioni per il recupero informativo di documentazione del settore. In appendice viene data una rappresentazione tabellare (Tabelle 1, 2 e 3) dei termini estratti dai titoli degli articoli delle Conferenze Internazionali di Linguistica Computazionale¹ (1965-2010) da cui si evince la rilevanza dei temi argomentali propria di questo settore disciplinare. Il sistema usato, per l'elaborazione di tali dati, è disponibile presso l'Istituto di Linguistica Computazionale "A. Zampolli" del CNR di Pisa.

Keywords

Linguistica computazionale, Informatica, Terminologia, Estrazione dell'informazione.

1. Accostamento elettronica-linguistica

La storia dell'uomo è caratterizzata dalla ricerca di tecniche di rappresentazione dell'informazione in supporti diversi che transitano dalle pietre alle tavolette di argilla, ai papiri, alle pergamene e alla carta. Negli anni '50 nacque un nuovo supporto, quello elettronico. L'informazione iniziò così a essere rappresentata e manipolata dai primi calcolatori elettronici per i vantaggi che offrivano nell'analisi di un testo scritto². Non solo,

¹ COLING, International Conference on Computational Linguistics. Le conferenze del COLING sono organizzate dal Comitato internazionale di Linguistica Computazionale /International Committee on Computational Linguistics (ICCL). La prima conferenza è stata organizzata a New York nel 1965.

² Padre Roberto Busa fu il pioniere della lessicografia computazionale iniziando nel 1951 la compilazione automatica delle concordanze dell'opera di Tommaso d'Aquino, per un totale di 10 milioni di occorrenze presso il Centro per l'Automazione dell'Analisi Linguistica (CAAL) di Gallarate, creato principalmente con finanziamenti della IBM Italia. Cfr. ROBERTO BUSA, *Varia Specimina Concordantiarum : Sancti Thomae*

anche i pioneristici tentativi di traduzione automatica³ (TA) appartennero a questo periodo, tutti incentrati nella formalizzazione di grammatiche che simulano al calcolatore la comprensione del linguaggio umano, per generare frasi sintatticamente corrette per la traduzione nelle lingue naturali. Nella realtà, la traduzione automatica all'inizio fu sostanzialmente una mera traduzione di parole perché i processi sintattici all'epoca erano ancora troppo poco sviluppati per poter riprodurre automaticamente il senso di un testo scritto. Nel 1954 ebbe luogo la prima esperienza di un sistema di traduzione automatica alla Georgetown University negli Stati Uniti. Il programma sviluppato dall'IBM, capace di tradurre frasi dal russo verso l'inglese, consisteva in un vocabolario di 250 parole e sei regole di grammatica. Pioniere del settore della Traduzione Automatica e delle possibili applicazioni del computer alla scienza linguistica fu Yehoshua Bar-Hillel⁴, ricercatore del MIT Research Laboratory of Electronics.

Di pari passo alla TA si avviò anche il trattamento elettronico dei testi che consiste in una serie di tecniche e procedure di spoglio elettronico, applicate ad opere letterarie per lo studio del lessico di una lingua. Queste elaborazioni di dati linguistici permisero la creazione di indici, di concordanze, di analisi quantitative sulle frequenze delle parole, interessando settori quali la lessicografia, la stilistica, la critica dei testi, la statistica linguistica, la dialettologia, la gestione degli archivi e delle biblioteche, la psicologia e la psichiatria del linguaggio.

L'elettronica è stata utilizzata fin da subito per la rappresentazione di dati linguistici per la grande facoltà di manipolazione dell'informazione testuale. Mai nella storia dell'uomo un supporto si era reso tanto flessibile: i testi in formato elettronico, per gli inglesi *Machine-readable form*, potevano finalmente essere riprodotti *ad infinitum* e corretti, le parole che lo componevano potevano essere rimpiazzate, ordinate alfabeticamente e contestualizzate, era una sorta di magia. I supporti cartacei non avrebbero mai permesso tanti possibili cambiamenti in un testo scritto. A tal proposito il Professor Antonio Zampolli⁵, scrive: "Dall'inizio degli anni '40 la stampa ha spesso commentato in modo suggestivo

Aquinatis Hymnorum Ritualium, Archivum Philosophicum Aloisianum, a cura della facoltà di Filosofia dell'Istituto Aloisiaum S.J. ; Serie II - N. 7. Milano : Fratelli Bocca Editori, 1951.

³ Atto ufficiale della nascita della traduzione automatica è il memorandum *Translation* di Warren Weaver. Cfr. Repr. in: Locke, W.N. and Booth, A.D. (eds.) *Machine translation of languages: fourteen essays* (Cambridge, Mass.: Technology Press of the Massachusetts Institute of Technology, 1955).

⁴ YEHOOSHUA BAR-HILLEL (1951): *The present state of research on mechanical translation*, *American Documentation* 2 (4), 229-237. Reprinted (with amendments) as 'The state of machine translation in 1951' in: Y.Bar-Hillel, *Language and information* (Reading, Mass.: Addison-Wesley, 1964), 153-165.

⁵ Antonio Zampolli è il fondatore della Sezione di Linguistica presso l'Istituto CNUCE del CNR di Pisa, nel 1968 trasformata successivamente in Laboratorio di Linguistica Computazionale, per poi diventare nel 1978

l'accostamento elettronica-linguistica. Allora come oggi l'adozione di mezzi d'avanguardia non giustifica preoccupazione o esaltazione alcuna per l'incontro tecnica-umanesimo: il calcolatore si è rivelato ancora una volta un potente strumento al servizio del lavoro umano"⁶. I calcolatori al servizio del lavoro umano vennero sempre più utilizzati anche per la costruzione di banche dati, contenitori elettronici di informazione, in grado di recuperare elementi pertinenti a domande poste dagli utilizzatori per mezzo di un software dedicato. In particolare, parliamo delle banche dati testuali, costituite da grandi collezioni di documenti per i vari settori del sapere. Un esempio autorevole in Italia negli anni '60 è fornito dagli spogli elettronici per la costituzione dell'archivio lessicale⁷ dell'Accademia della Crusca per la redazione del vocabolario storico della lingua italiana con il contributo del Consiglio Nazionale delle Ricerche, presso l'Istituto CNUCE di Pisa. Sempre in Italia nel 1967 l'Accademia dei Lincei organizzò un Convegno dal titolo "L'Automazione elettronica e le sue implicazioni scientifiche, tecniche e sociali"⁸.

Da allora la costruzione di banche di dati linguistici, prototipi di biblioteche elettroniche, conservate su nastro magnetico nei vari Centri⁹ europei non si è mai interrotta, arrivando dritta all'era digitale, non più su nastro ma direttamente in rete, la *Gallica* e *Google Libri*, sono solo alcuni esempi .

Queste nuove risorse elettroniche del sapere, questi grandi corpora, questi *repository* si aggregano paradossalmente per andare a formare un unico grande corpus, che può essere osservato, indagato e studiato, prima ancora di essere interrogato per ottenere informazioni pertinenti : "The web is immense, free and available by mouse-click. It contains hundreds of

Istituto di Linguistica Computazionale del Consiglio Nazionale delle Ricerche. Il Professor Antonio Zampolli ha diretto l'Istituto fino al 2003. Oggi l'Istituto è intitolato alla sua memoria.

⁶ ANTONIO ZAMPOLLI, *L'elaboratore elettronico negli studi linguistici*, Estratto dalla rivista IBM n. 2, maggio 1968.

⁷ Cfr. ALDO DURO, *Élaborations électroniques de textes effectuées par l'Accademia della Crusca, pour la préparation du dictionnaire historique de la langue italienne*, *Linguistica Matematica e Calcolatori*, Atti del Convegno e della prima scuola internazionale, a cura di Antonio Zampolli, Firenze Leo S. Olschki editore , 1973.pagine 53-75.

⁸ Cfr. *L'automazione elettronica e le sue implicazioni scientifiche, tecniche e sociali*, *Problemi attuali di Scienza e di Cultura*, Roma, 16-19 ottobre 1967. Accademia Nazionale dei Lincei, Anno CCCLXV - 1968, Quaderno n. 110.

⁹ Ricordiamo tra i tanti Centri europei creati a cavallo degli anni '50 e '60: le Laboratoire d'analyse lexicologique di Besançon, le Centre de linguistique automatique appliquée di Bruxelles, il Centro per l'automazione dell'analisi linguistica di Gallarate, le laboratoire d'analyse statistique des langues anciennes de l'Université de Liège , le Centre de recherche pour un trésor de la langue française di Nancy, the K.V.A.L. group research for quantitative linguistics di Stoccolma, le centre des recherches sur la syntaxe de l'allemand di Saarbrücken.

billions of words of text and can be used for all manner of language research” (Kilgarriff and Grefenstette, 2003)¹⁰.

I miliardi di parole della lingua naturale usata dai fruitori del WEB appartengono a diversi sotto-linguaggi destinati a usi ben precisi, a domini ristretti e tecnici e delegati a diventare nuclei di conoscenza scientifica nei moderni canali del sapere, grandi oceani informativi che caratterizzano il nostro tempo.

In questo scenario il linguaggio usato veicola il pensiero e i termini possono esprimere e orientare la nostra visione del mondo. Lo sviluppo del lessico negli anni diventa la testimonianza di un bisogno comunicativo mai soddisfatto. In tutti i settori, le parole possono diventare ancora di recupero informativo per andare a estrarre porzioni di conoscenza dai flussi informativi in rete.

“I termini sono il motore di un processo di comprensione, nella misura in cui essi rilevano una comprensione nuova e una comprensione vecchia, tanto che per la terminologia classica, i termini denotano i concetti”, ha scritto nel 2000 Rita Temmerman, docente di Linguistica Applicata presso Erasmushogeschool di Brussel¹¹.

2. *Necessità di nominazione*

“Since its origins, Computational Linguistics has been in an intermediate position between Computer Science and Artificial Intelligence, Linguistics and Cognitive Science, and Engineering. Computer Science itself shares its roots with Computational Linguistics”¹², Giacomo Ferrari, Preside della Facoltà di Lettere e Filosofia dell’Università degli Studi del Piemonte Orientale, in maniera concisa e chiara, delinea l’interdisciplinarietà della linguistica computazionale. Questa interdisciplinarietà è espressa dal contenuto argomentale delle ricerche del settore che, fin dagli albori, rispondono a discipline come l’informatica, la linguistica, la linguistica applicata, la linguistica matematica, la linguistica statistica e la linguistica algebrica. *Linguistica Computazionale*. L’aggettivo ‘computazionale’ svela il legame diretto con il sostantivo computer. A partire dagli anni ’60 furono usati differenti

¹⁰ Introduction to the Special Issue on the Web as Corpus. *Computational Linguistics*, Volume 29, 3, Association for Computational Linguistics, 333-347.

¹¹RITA TIMMERMAN, *Towards New Ways of Terminology Description. The sociocognitive approach*, Amsterdam/Philadelphie, John Benjamins, 2000, pag. 228.

¹²GIACOMO FERRARI, A State of the Art in Computational Linguistics, in P.van Sterkenburg (ed.), *Linguistics Today - Facing a Great Challenge*, John Benjamins, 2003, pag. 163.

termini per i giovani filoni di ricerca, ci fu necessità di nominazione, di traduzione e di ricorso al neologismo. Prendiamo ad esempio il caso francese: Linguistica Computazionale, viene tradotto in un primo periodo con “Linguistique Computationnelle”, nel 1992 viene suggerito il termine “Linguistique Informatique”. Riportiamo le parole della prefazione agli atti della 15th International Conference on Computational Linguistics, COLING 1992, del Professor Christian Boitet, Professeur à l'Université Joseph Fourier (Grenoble 1) “D'aucuns ont remarqué que, dans les versions françaises des diverses annonces de COLING-92, nous avons remplacé l'ancien terme "linguistique computationnelle" par "linguistique informatique". Bien sûr l'adjectif "computationnel" sonne très mal en français, même s'il est régulièrement formé. Mais surtout, il semble limiter notre domaine à l'étude des processus algorithmiques de traitement effectif de textes”¹³. Già nel 1991 in un testo spagnolo si legge: “El campo científico correspondiente se denomina con frecuencia Informática lingüística o lingüística informática, por lo general, en el dominio de la Ciencia Cognitiva, en la intersección de la inteligencia artificial, de la lingüística, de la lógica, de la psicología y de la filosofía.”¹⁴

Quasi quaranta anni prima, nel 1955, il Professor Jacques Perret, professore di filologia latina alla Sorbonne propose in una lettera,¹⁵ diventata famosa, la parola “ordinateur”, anziché “calculateur” per riprodurre la parola inglese “computer”, ancora nel '69 il termine francese “calculateur” veniva utilizzato ai convegni e nelle pubblicazioni scientifiche.¹⁶ Anche il termine “informatica” è un neologismo, l'esigenza di nominazione si estese anche al termine inglese “computer science”, occorre una riflessione sulle possibilità di una buona traduzione. Philippe Dreyfus, ingegnere e direttore del Centre National de Calcul Électronique alla Bull, finalmente nel 1962 inventa il termine “informatique”, elaborato dalla contrazione di due parole “information” e “automatique”.¹⁷ Il termine fu esportato in Italia, Germania, Spagna, Portogallo e Russia.

¹³ CHRISTIAN BOITET, À propos de ces actes – About these proceedings, Actes de COLING-92, Nantes, 1992, p. 10.

¹⁴ GÉRARD SABAH, El tratamiento automático del lenguaje escrito, in José Vidal Beneyto, *Las Industrias de la Lengua*, Madrid, Pirámide, 1991, pag. 41.

¹⁵ ORDINATEUR: la lettre de Jacques Perret . <http://www.les-infostrateges.com/article/0505281/ordinateur-la-lettre-de-jperret>

¹⁶ Colloque international sur l'emploi des calculateurs en archéologie : problèmes sémiologiques et mathématiques, a cura di Jean-Claude Gardin, Marseille, C.N.R.S., 1969.

¹⁷ INFORMATIQUE. Néologisme construit à partir des mots information et automatique par P.Dreyfus en 1962. Il s'agit donc d'une discipline qui concerne le traitement automatique de l'information. La définition acceptée par l'Académie Française est la suivante : "science du traitement rationnel, notamment par machines automatiques, de l'information considérée comme le support des connaissances humaines et des communication dans les

Negli anni '50 e '60 per designare operazioni computazionali nel settore della scienza del linguaggio vennero utilizzati volentieri aggettivi quali automatico e meccanico, e sostantivi quali meccanizzazione, macchina/machine a testimoniare un processo guidato da congegni elettronici, in cui l'uomo non è il principale attore. Questi aggettivi sono abbinati sovente al sostantivo traduzione, mentre i rispettivi sostantivi "mechanization" e automazione sono applicati per indicare il complesso delle operazioni in quel settore.

A titolo esemplificativo si riportano alcuni titoli di articoli e di libri scritti a cavallo degli anni '50 e '60 dagli addetti al settore:

*A survey of Soviet work on automatic translation*¹⁸; *Mechanical Resolution of Linguistic Problems*¹⁹; *On the Problem of Mechanical Translation*²⁰; *La Machine à Traduire*²¹; *La Mécanisation dans les Recherches Lexicologiques*²²; *Linguistic Analysis and Programming for Mechanical Translation*²³; *Les travaux lexicologiques préparatoires à la traduction automatique*²⁴; *Readings in Automatic Language Processing*²⁵; *Les Machines dans la Linguistique*²⁶; *Machine Translation*²⁷; *Automated Language Processing*²⁸.

Nel 1966 appare per la prima volta, usato in una fonte ufficiale, il termine "Computational Linguistics" (CL), nel famoso *ALPAC Report*²⁹ dell'*Automatic Language Processing Advisory Committee* sulla traduzione automatica. Fino ad allora coppie di termini e gruppi di parole come "Automatic Language Processing", "Computation in Linguistics", "Mechanical

domaines techniques, économiques et sociaux", *Dictionnaire de l' informatique et d'internet*, Dicofr.com. <http://www.dicofr.com/cgi-bin/n.pl/dicofr/definition/20010101002821>

¹⁸ Cfr. ANTHONY G. OETTINGER (1958). A survey of Soviet work on automatic translation. *Mechanical Translation* 5(3).101-110.

¹⁹ Cfr. A. DONALD BOOTH, (1965), *Mechanical resolution of linguistic problems*. In: Kent, A. & Taulbee, O.E. eds. *Electronic*

information handling (Washington, D.C.: Spartan Books), 41-50. <http://mt-archive.info/Kent-1965-Booth.pdf>

²⁰ Cfr. DMITRI PANOV (1956), *The Problem of Mechanical Translation*, *Mechanical Translation*, vol.3, no.2, November 1956, pp. 42-43.

²¹ Cfr. ÉMILE DELAVENAY (1959), *La machine à Traduire* Delavenay, Paris, Presses Universitaire de France.

²² Cfr. Actes du Colloque International sur la Mécanisation des Recherches Lexicologiques (1962), *Cahiers de Lexicologie*, Vol. 3, 62, Publiés par Bernard Quemada, Paris, Didier - Larousse.

²³ Cfr. SILVIO CECCATO (1961), *Linguistic analysis and programming for mechanical translation*, Gordon & Breach, New York, 1961.

²⁴ Cfr. BERNARD POTTIER (1962), *Les travaux lexicologiques préparatoires à la traduction automatique*, *Cahiers de Lexicologie*, Vol. 3, 62, Publiés par Bernard Quemada, Paris, Didier - Larousse, pagine 200-206.

²⁵ Cfr. DAVID G. HAYS (1966) *Readings in Automatic Language Processing*, New York, American Elsevier Pub. Co.

²⁶ Cfr. JITKA STINDLOVÁ, ERICH MATER (Rédaction), (1968),. *Les machines dans la linguistique: colloque international sur la mécanisation et l'automation des recherches linguistiques*, Prague. Academie Techecoslovaque des Sciences et l'Academie Allemande des Sciences.

²⁷ Cfr. ANDREW DONALD BOOTH (ed) (1967), *Machine translation*, Amsterdam, North-Holland.

²⁸ Cfr. HAROLD BORKO (1967), *Automated language processing*, New York, Wiley, 1967.

²⁹ ALPAC (1966), *Language and machines: computers in translation and linguistics*. A report by the Automatic Language Processing Advisory Committee. National Academy of Sciences, Washington, DC.

Translation” erano utilizzati pressappoco con lo stesso significato per indicare l’automazione nelle ricerche linguistiche.

Negli anni '60 le ricerche del trattamento automatico della lingua si consolidano per dare vita ad associazioni di settore e alla creazione di riviste internazionali che riflettevano studi, tendenze e applicazioni della LC e della TA. Data la vastità della documentazione in proposito, ne ricordiamo solo alcune:

- Nel 1959 nasce l’*Association pour l’étude et le développement de la Traduction Automatique et de la Linguistique Appliquée* (ATALA) che nel 1965 diventa l’*Association pour le Traitement Automatique des Langues*];
- 1962: nasce *The Association for Machine Translation and Computational Linguistics* (AMTCL) che nel 1968 diventa *The Association for Computational Linguistics* (ACL);
- 1965: nasce l’*Association Internationale de Linguistique Appliquée o International Association of Applied Linguistics* (AILA);
- 1973: nasce l’*Association for Literary and Linguistic Computing* (ALLC);
- 1978: nasce l’*Association for Computers and the Humanities* (ACH).
- Nel 1954 nasce la rivista *Mechanical Translation* (MT), pubblicata dal Massachusetts Institute of Technology che nel 1965 cambia titolo in *Mechanical Translation and Computational Linguistics* (AMTCL);
- 1960: nasce la rivista *Cahiers de Lexicologie*;
- 1960: nasce la rivista *La Traduction Automatique* che nel 1965 cambia nome in *TA Informations, Revue internationale du traitement automatique du langage*;
- 1966: nasce la rivista *Computers and the Humanities* che con il primo numero del 2006 cambia titolo in *Language Resources and Evaluation*;
- 1974: nasce l’*American Journal of Computational Linguistics* (AJCL) che cambia nome nel 1984 in *Computational Linguistics*. Dal Volume 35, fascicolo 1, del 2009 la rivista diventa open access e non viene più stampata.
- 1981: nasce la rivista *Linguistica Computazionale* pubblicata a cura dell’Istituto di Linguistica Computazionale del CNR di Pisa.

Dal 1965, anno della prima conferenza di Linguistica Computazionale COLING, tenuta a New York, l’elettronica e la linguistica hanno percorso insieme molta strada, sono partite

dalle concordanze di San Tommaso d'Aquino di Padre Busa per arrivare fino alla creazione di indici e concordanze dei motori di ricerca per il recupero informativo e la creazione di ontologie computazionali per l'estrazione intelligente di informazione. “Gli sviluppi più recenti della linguistica computazionale e del *natural language engineering* hanno creato soluzioni tecnologiche dalle enormi potenzialità per migliorare la ricerca e gestione intelligente dell'informazione contenuta nei documenti testuali. Le nuove tecnologie della lingua, infatti, permettono ai sistemi informatici di accedere al contenuto digitale attraverso il Trattamento Automatico della Lingua (TAL) o Natural Language Processing (NLP)”.³⁰

Appendice

Il Corpus COLING, composto di circa 50.000 parole dei titoli delle 23 conferenze fondamentali per la storia della Linguistica Computazionale, è stato utilizzato per effettuare una indagine terminologica sulla rilevanza di temi argomentali del settore. Il sistema usato, per l'elaborazione di tali dati, è disponibile presso l'Istituto di Linguistica Computazionale del CNR di Pisa, laboratorio Dylan Lab.³¹ Trattasi di una piattaforma software di supporto avanzato alla gestione documentale per la creazione dinamica di repertori terminologici e ontologie di dominio a partire da testi, anche non strutturati, e per l'indicizzazione concettuale di documenti. Nel nostro caso l'analisi del testo ha generato tre tipologie di risultati basati sulla rilevanza percentuale di termini multipli, di termini di domini specifici e di termini singoli come è visibile nelle Tabelle 1, 2 e 3. Interessante è notare come il tema argomentale “machine translation” nell'intero Corpus abbia una rilevanza del 100% e tematiche come “sense disambiguation” abbiano, nella tabella 2, “domain-specific multiple terms”, una rilevanza del 100%.

³⁰ NICOLETTA CALZOLARI, ALESSANDRO LENCI (2004), *Mondo Digitale*, numero 2, giugno.

³¹ DYLAN LAB, *Laboratorio di modelli computazionali delle dinamiche del linguaggio e della cognizione* dell'Istituto di Linguistica Computazionale del CNR di Pisa.

Tabella 1





















ID	relevant multiple terms (general purpose relevance)	relevance
1	machine translation	100.0% 
2	sense disambiguation	21.1% 
3	word sense	13.6% 
4	word sense disambiguation	12.8% 
5	example-based machine translation	7.6% 
6	japanese sentence	6.3% 
7	unknown word	5.8% 
8	language translation	4.9% 
9	chinese character	4.0% 
10	sentiment classification	4.0% 
11	knowledge-based machine translation	3.8% 
12	parallel text	3.2% 
13	analysis of japanese compound	3.2% 
14	chinese word sense disambiguation	3.1% 
15	example-based machine	2.6% 
16	pattern-based machine translation	2.5% 
17	machine translation evaluation	2.5% 
18	natural language access	2.5% 
19	japanese-to-english machine translation	2.5% 
20	chinese sentence	2.4% 

Tabella 2











































ID	domain-specific multiple terms (text-dependent relevance)	relevance
1	machine translation	100.0% 
2	sense disambiguation	100.0% 
3	word sense	100.0% 
4	word sense disambiguation	100.0% 
5	japanese sentence	100.0% 
6	unknown word	100.0% 
7	example-based machine translation	100.0% 
8	language translation	100.0% 
9	sentiment classification	100.0% 
10	example-based machine	100.0% 
11	parallel text	100.0% 
12	chinese character	100.0% 
13	knowledge-based machine translation	100.0% 
14	chinese sentence	100.0% 
15	analysis of japanese compound	100.0% 
16	chinese word sense disambiguation	100.0% 
15	measuring	72.4% 
16	classification	72.3% 
17	verbs	72.0% 
18	framework	70.5% 
19	grammars	70.0% 
20	language	69.2% 

Tabella 3

ID	single terms	relevance
1	translation	100.0% 
2	sentence	92.0% 
3	disambiguation	91.0% 
4	improving	89.4% 
5	machine	85.5% 
6	application	85.3% 
7	sense	83.1% 
8	transfer	80.4% 
9	evaluation	79.6% 
10	word	79.2% 
11	text	78.7% 
12	pattern	77.4% 
13	simple	76.2% 
14	analysis	73.7% 
15	measuring	72.4% 
16	classification	72.3% 
17	verbs	72.0% 
18	framework	70.5% 
19	grammars	70.0% 
20	language	69.2% 

LINK

Biblioteca ILC. Visibile dal Metaopac del Sistema Bibliotecario Pisano:

<http://leonardo.isti.cnr.it/Isis/servlet/Isis?Conf=../Conf/AIXnotlogic0.pisa.sys.file>

Pubblicazioni dell'Istituto di Linguistica Computazionale "Antonio Zampolli". Visibili dal Sistema PuMa Publication Management dell'Istituto ISTI dell'Area della Ricerca di Pisa:

<http://puma.isti.cnr.it//index.php?langver=it>

Catalogo per soggetti. Visibile dal sito del Progetto BIBLOS, Biblioteca virtuale del settore umanistico del Consiglio Nazionale delle Ricerche:

http://www.biblos.cnr.it/04_8_4.html

Fondo Antonio Zampolli - Bibliografia sul Trattamento Automatico della Lingua:

HTTP://WWW.BIBLOS.CNR.IT/04_2_TALB.HTML

ACL Anthology : A Digital Archive of Research Papers in Computational Linguistics:

<http://aclweb.org/anthology-new/>

Machine Translation Archive, Electronic repository and bibliography of articles, books and papers on topics, in machine translation, computer translation systems, and computer-based translation tools dell' *European Association for Machine Translation EAMT*:

<http://www.mt-archive.info/>

DyLan Lab: Laboratorio di modelli computazionali delle dinamiche del linguaggio e della cognizione

<http://www.ilc.cnr.it/viewpage.php/sez=ricerca/id=1008/vers=ita>

<http://www.ilc.cnr.it/dylanlab/>