IRIS AperTO

UNIVERSITÀ DEGLI STUDI DI TORINO

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

Live Chat e Digital Humanities: il Reference Online incontra il Text Mining

This is a pre print version of the following article:

Original Citation:

Availability:

This version is available http://hdl.handle.net/2318/1662730 since 2018-05-23T14:53:34Z

Terms of use:

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)

06 October 2026

Live Chat e Digital Humanities: il Reference Online incontra il Text Mining.

Maria Vittoria Muzzupapa - Marco Stefano Tomatis

Abstract

Dal 2014 il Sistema Bibliotecario di Ateneo di Torino ha attivato una live chat, un servizio di reference online sincrono, per assistere i propri utenti nelle attività di ricerca bibliografica e di accesso alle risorse digitali. Tale servizio è stato gestito e coordinato dalla Sezione Servizi Bibliografici Digitali, i cui bibliotecari hanno sostenuto in questi 3 anni un totale di 3381 conversazioni. Il crescente successo del servizio ha portato lo staff responsabile del progetto a valutare queste stesse conversazioni come potenziali fonti di informazioni utili a far emergere le problematiche espresse dagli utenti. Si è deciso quindi di unire due professionalità, quelle del bibliotecario e del linguista computazionale, per analizzare gli scambi operatore/utente. Tramite tecniche di text mining sono state estratte informazioni di tipo quantitativo e qualitativo sull'attività di reference. Proprio a partire dai risultati di queste analisi è stato ad esempio possibile produrre una lista di FAQ[1] e selezionare le richieste rivolte a una singola biblioteca[2], con lo scopo di offrire un ulteriore miglioramento dei servizi di reference.

Parallelamente alle attività menzionate, è stato altresì possibile osservare i diversi vocabolari e stili di linguaggio degli utenti e degli operatori della chat. Allo scopo di migliorare l'interazione utente/bibliotecario durante l'intervista di reference, è stata analizzata con metodologie proprie delle digital humanities la diversa terminologia usata dai due interlocutori. In particolare l'attenzione si è concentrata sulle differenti espressioni usate da entrambi per definire i servizi bibliotecari, gli strumenti per la ricerca e le tecnologie offerti.

Premesse

Il reference digitale è ormai acquisito come naturale mezzo di comunicazione con gli utenti delle biblioteche: sono diffusi i sistemi di *helpdesk,* le comunicazioni via mail e anche la chat ormai è un canale ampiamente conosciuto. Proprio su quest'ultimo mezzo vuole concentrarsi lo studio che andremo a esporre, in particolare partendo dall'esperienza realizzatasi presso il Sistema Bibliotecario dell'Ateneo di Torino (da qui in poi SBA), che ha attivato il servizio di live chat sul

[1] Maria Vittoria Muzzupapa, Marco Stefano Tomatis, e Franco Carlo Bungaro, «Live chat e natural language processing in sinergia per il miglioramento dei servizi bibliotecari.», *Biblioteche oggi Trends* 3, n. 1 (2017): 41–53, https://doi.org/10.3302/2421-3810-201701-041-1.

[2] Franco Bungaro, Maria Muzzupapa, e Marco Tomatis, «Extending the Live Chat Reference Service at the University of Turin – A Case Study», *Proceedings of the IATUL Conferences*, 2017, https://docs.lib.purdue.edu/iatul/2017/challenges/1.

proprio sito istituzionale e su quello del discovery tool d'Ateneo[3]. Ciò che si vuole analizzare in profondità, basandosi su esempi pratici, cioè su estratti di conversazioni, è come si articolano gli scambi e in che modo utente e bibliotecario di reference dialogano e si passano informazioni.
SnapEngage[4], il software con cui è realizzata la chat usata dal SBA, salva tutte le conversazioni in file di log in formato testo. Questo ha permesso la raccolta di moltissimi dati, consistenti in tutti i contenuti dei dialoghi avvenuti tra utenti e bibliotecari: le domande e risposte, le reazioni, le problematiche degli utenti. Ogni singolo vocabolo componente gli scambi è diventato un'informazione analizzabile e interpretabile.
Sappiamo che buona parte della complessità di un'attività di reference in chat è data dalla mancanza di un rapporto vis-à-vis, dall'impossibilità di comunicare con gestualità, toni della voce ed espressioni del volto [5]:

> Verbal communication is the primary method of information exchange between librarians and users in face-to-face reference encounters, where facial expressions and body language are additional cues in the communication process. In the chat reference situation, no visual or audio cues are available and librarians can only communicate with users by the exchange of written message.

IFLA[6] e RUSA[7] si sono impegnate a stilare linee guida utili alla gestione del reference virtuale, compreso quello sincrono. Entrambe le istituzioni fanno un chiaro riferimento alla necessità di buone doti di comunicazione e invitano a una gestione della conversazione naturale e colloquiale.
Molte delle istruzioni e delle analisi prodotte dalla letteratura di settore, dalle istituzioni accademiche bibliotecarie che si sono dotate della live chat approfondiscono la metodologia di conduzione delle interviste di reference con questo mezzo e, quindi, la gestione dell'attività di information retrieval e di passaggio delle informazioni durante una sessione di assistenza interamente virtuale. Particolarmente significativa è l'accurata analisi operata da Lankes nel 2009[8].
All'interno di questo panorama, il contributo della ricerca qui proposta vuole essere un'analisi del materiale collezionato in quasi 4 anni di attività di reference in chat, esaminando i contenuti delle conversazioni tra bibliotecari e utenti. In modo particolare si sono volute studiare le differenze espressive e le dinamiche di e tra gli interlocutori partendo dall'osservazione delle parole usate da entrambi.

---

[3] www.tutto.unito.it
[4] Per maggiori informazioni su SnapEngage: https://snapengage.com/
[5] Lili Luo, «Chat reference competencies: identification from a literature review and librarian interviews», *Reference Services Review* 35, n. 2 (2007): 195–209, https://doi.org/10.1108/00907320710749137.
[6] IFLA, «IFLA Digital Reference Guidelines», 2002, https://www.ifla.org/files/assets/reference-and-information-services/publications/ifla-digital-reference-guidelines-en.pdf.
[7] RUSA, «Guidelines for Implementing and Maintaining Virtual Reference Services», Text, Reference & User Services Association (RUSA), 2017, http://www.ala.org/rusa/resources/guidelines/virtrefguidelines.
[8] R. David Lankes, «New Concepts in Digital Reference», *Synthesis Lectures on Information Concepts, Retrieval, and Services* 1, n. 1 (2009), http://www.morganclaypool.com/doi/abs/10.2200/S00166ED1V01Y200812ICR001.

Il linguaggio

Il linguaggio verbale, in questo caso scritto, può essere un elemento ricchissimo di informazioni, capaci di fornire numerosi spunti di analisi sul modo in cui gli utenti comunicano e i bibliotecari di reference comprendono e reagiscono alle richieste.

Nel nostro caso i due soggetti condividono il contesto ambientale dell'Università di Torino e nello specifico l'ambito operativo delle attività, cioè il sistema bibliotecario. Teoricamente dovrebbero quindi muoversi su un terreno lessicale comune.

L'analisi semiautomatica eseguita sui file di log ha fatto emergere termini ricorrenti che non solo hanno portato all'individuazione di criticità sollevate dagli utenti, ma hanno anche sottolineato l'esistenza di un vocabolario caratterizzante il servizio stesso di live chat. La terminologia impiegata da entrambi gli interlocutori è ovviamente determinata dal contesto: trattandosi di servizi bibliotecari ricorrono le parole "libro", "ricerca", "catalogo", "biblioteca", "articolo".

Cominciamo le nostre osservazioni partendo dalla parola "biblioteca". Durante le chat di reference è emerso spesso come questa parola indichi per il bibliotecario il luogo specifico in cui l'utente deve recarsi per ottenere un determinato servizio e per reperire informazioni specifiche o la copia di un documento richiesto. In altri casi invece, si è visto come l'utente abbia esteso il concetto di biblioteca dall'ambiente fisico a quello digitale, spesso associando il luogo di proprio riferimento per il reperimento di documenti alle attività che deve eseguire online, quasi che la biblioteca fisica rappresenti anche in un ambiente digitale, e quindi virtuale, l'insieme delle attività e dei servizi per la ricerca bibliografica[9].

Visitor : 10:12:47: ciao , grazie mille ! come faccio ad accedere alla biblioteca Bobbio ? Ho già impostato tutto , ma non capisco da dove posso accedere ..
Maria Vittoria : 10:13:11: http://www.bibliotecabobbio.unito.it/it
Maria Vittoria : 10:13:19: questo è il sito della bobbio
Maria Vittoria : 10:13:29: cosa hai già impostato ?
Maria Vittoria : 10:13:37: il proxy ?
Visitor : 10:13:39: il proxy unito
Maria Vittoria: 10:13:41: ok
Visitor: 10:13:50: grazie
Maria Vittoria: 10:14:06: se stai cercando una banca dati in particolare
Maria Vittoria: 10:14:13: e hai problemi
Maria Vittoria: 10:14:17: posso provare ad aiutarti?
Visitor: 10:15:14: no, sto cercando alcuni manuali, ma temo non ci siano proprio... grazie mille..
Maria Vittoria: 10:15:24: ok
Maria Vittoria: 10:15:33: se hai hai bisogno fammi sapere
Visitor: 10:19:25: grazie davvero!

Nell'esempio di conversazione sopra proposto, la richiesta iniziale sembra essere il bisogno di "accedere" alla struttura Bobbio (intende il luogo fisico o il sito web?), per poi risolversi

[9] Se dovessimo paragonare questo fenomeno a una figura retorica, potremmo accostarlo alla sineddoche, che con la parte rappresenta il tutto.

apparentemente con l'accesso da casa tramite proxy[10] alle risorse elettroniche sottoscritte da UniTO. Infine l'utente svela l'intento di cercare dei manuali, senza però chiarire se elettronici o cartacei, e quindi mostra una certa confusione tra ambiente fisico e ambiente digitale.

Da questo esempio, si comprende come l'ambientazione digitalizzata della conversazione influisca profondamente sui contenuti dei dialoghi e forse anche sulle aspettative di chi cerca, che proprio perché sta agendo nel web, si aspetta di trovare tutti i documenti in formato digitale. Un'assistenza bibliografica in live chat aggiunge all'intervista di reference una serie di problematiche che nella conversazione in presenza non si verificano. Le complicazioni non sono provocate solo dall'impossibilità della comunicazione attraverso sguardi e gesti, ma anche dalla smaterializzazione degli spazi.

Tutto ciò obbliga chi fa assistenza ad avere una capacità interpretativa, soprattutto inferenziale. L'ambiente digitale annulla i confini entro cui la conversazione avviene e moltiplica la possibili interpretazioni di un'affermazione o di una richiesta, soprattutto per quanto riguarda la natura (fisica o virtuale) di ciò di cui si sta parlando.

L'utente che usa la chat sta navigando, sta facendo ricerca online e proprio in questa dimensione sta contestualizzando tutte le sue attività. Ciò vuol dire che per chiedere aiuto avrà bisogno di un vocabolario arricchito di termini specifici e legati alla tecnologia a cui ricorre e che esprimano in maniera tecnica le sue necessità.

Entrano pertanto in gioco nella nostra analisi parole quali "sito", "accesso", "proxy", "collegarsi".

Abbiamo visto poco sopra come spesso gli utenti non siano realmente consapevoli di quello che devono fare e come. "Proxy" per esempio è un vocabolo che talvolta è acquisito, ma le azioni che sono dietro la configurazione del servizio a cui si riferisce non sempre sono ben comprese. Si sa magari che il proxy serve a vedere le risorse delle biblioteche accademiche da casa, ma non si sa tanto in che modo funziona. Ciò comporta che alle volte sia usato dagli utenti in maniera non troppo appropriata all'interno della conversazione.

Visitor: 10:30:49: Mi sono registrata per poter accedere a proxy da casa

Il proxy è un servizio che si configura sul browser con cui si sta navigando (Firefox, Chrome, Edge, etc.) e non necessita di una registrazione preliminare.

Tuttavia, le difficoltà espressive sono emerse anche su termini apparentemente più semplici e di maggiore utilizzo nel quotidiano. Parliamo di parole come "registrarsi" e "accedere".

Visitor : 12:19:40: ciao avrei bisogno di registrarmi per poter accedere ai dizionari online zanichelli
Maria Vittoria : 12:19:52: sì
Maria Vittoria : 12:20:04: registrarti su zanichelli o
Maria Vittoria : 12:20:09: accedere da casa ?

[10] Software che, modificando le impostazioni del browser, consente all'utente di accedere ai servizi degli editori come se navigasse all'interno della rete di ateneo.

Visitor : 12:20:24: accedere da casa

Qui l'utente usa il termine "registrarmi" per indicare la necessità di accedere dalla rete di casa a una risorsa in abbonamento: formalmente la possibilità di registrarsi sulla piattaforma Zanichelli è possibile, ma è finalizzata alla gestione del proprio spazio personale. Per il bibliotecario in questo caso il campanello di allarme è dato dall'azione espressa dalle parole "accedere ai dizionari online", che indica con ogni probabilità il desiderio di visualizzare i contenuti della risorsa cui si sta facendo riferimento e non di accedere a una funzionalità della piattaforma. Chiesta la conferma su quale sia la reale intenzione da parte dell'utente, la conferma arriva per l'accesso da remoto.

A questo punto è possibile introdurre una problematica relativa al linguaggio oggetto di questo studio.

Già Ranganthan parlava di "slealtà del catalogo", in quanto strumento realizzato usando un "linguaggio familiare, naturale […] parole ordinarie"; ma se queste parole non sono "morfologicamente o radicalmente artificiali", la sintassi lo è e lo è anche la semantica[11]. Ebbene, se gli utenti continuano ad avere difficoltà a familiarizzare con la lingua del catalogo[12], con la ricerca online moltiplicano in maniera esponenziale le complicazioni a cui vanno incontro.

Anche se oggi il rapporto con la tecnologia è diventato inevitabile e se tutti noi navighiamo quotidianamente in rete, non è scontato che vi sia una reale consapevolezza degli strumenti[13] che si stanno usando. Gli utenti di una biblioteca accademica, quindi, oltre a dover imparare a interpretare la punteggiatura e la terminologia con cui i documenti sono descritti in un OPAC, ora devono confrontarsi con parole di uso comune, ma dai significati non sempre chiari.

Abbiamo già visto come per noi bibliotecari sia ben distinto il concetto di "accesso" da quello di "registrarsi", cosa che non è altrettanto chiara per gli utenti. Scopriamo che anche il concetto e la funzione che sono dietro il catalogo online, non sono ben acquisiti.

Qui di seguito riportiamo una conversazione dove si evidenzia chiaramente una certa confusione terminologica da parte di chi chiede assistenza, e probabilmente anche una scarsa consapevolezza delle operazioni che esegue per leggere da casa i contenuti delle riviste:

Visitor : 08:08:43: Ciao ! Ormai da due giorni non riesco ad accedere al servizio del catalogo ne passando attraverso il link del bibliopass ne provando ad entrare dal sito unito ! mi chiedevo se fosse un problema del pc o se il sito è furoi uso momentaneamente
Maria Vittoria : 08:09:04: ciao :)
Maria Vittoria : 08:09:12: è dovuto a un guasto tecnico
Maria Vittoria : 08:09:20: che si è verificato lunedì
Maria Vittoria : 08:09:29: Bibliopass non è ancora tornato in funzione
Maria Vittoria : 08:09:36: mentre il Proxy UniTO sì

[11] S. R. Ranganathan, *Il servizio di reference*, trad. da Carlo Bianchini (Firenze: Le Lettere, 2010), 81.
[12] Agnese Galeffi, «Se il catalogo parlasse, lo capiremmo? Cinque assiomi della comunicazione catalografica», *AIB studi* 57, n. 2 (2017), https://doi.org/10.2426/aibstudi-11648.
[13] Con strumenti si intendono sia quelli fisici, quali il computer o gli smartphone, sia quelli virtuali, come i data base o i software.

Il problema era causato da malfunzionamenti dei servizi per l'accesso alle risorse elettroniche fuori dalla rete UniTO[14]. Tuttavia l'utente non riesce a definirlo chiaramente nella sua domanda di apertura: qual è il servizio a cui non riesce ad accedere? Nomina il catalogo, ma a cosa si riferisce? Al catalogo, al discovery tool d'Ateneo o all'OpenURL Resolver?[15]. Infine si evidenzia un nuovo termine chiave: "sito". L'utente si riferisce a un "sito unito" dove cerca di entrare. Presumibilmente ha cercato di entrare nel servizio di autenticazione MyUnito disponibile sul portale web dell'Ateneo, che però non è funzionale alla modifica delle impostazioni di navigazione, ma a usufruire dei servizi forniti da UniTO in maniera personalizzata (segreteria, mail, ecc.).
Un altro esempio:

Visitor : 10:30:18: per accreditarsi
Maria Vittoria : 10:30:28: a cosa ?
Visitor : 10:30:55: dove la pagina di login per studenti
Maria Vittoria : 10:31:26: www.unito.it
Maria Vittoria : 10:31:32: in alto a destra
Maria Vittoria : 10:31:40: fai click su Login
Visitor : 10:32:05: grazie pensavo che si potesse fare anche da qui ,
Maria Vittoria : 10:32:12: ma cosa vuoi fare ?
Maria Vittoria : 10:32:20: ricerca di articoli online ?
Maria Vittoria : 10:32:26: o usare i servizi di segreteria ?
Visitor : 10:32:49: cerco un libro
Maria Vittoria : 10:32:52: ok
Maria Vittoria : 10:32:57: allora è diverso
Maria Vittoria : 10:33:11: sei a casa o in UniTO ?

La richiesta di apertura è significativa per due motivi: innanzitutto per la "sintesi" dell'utente, che non inserisce neanche il punto interrogativo. Sembra quasi che non sia del tutto convinto della presenza di qualcuno dietro il servizio di chat. Più avanti svilupperemo meglio questo discorso.
Il secondo aspetto rilevante della frase è il termine "accreditarsi" che, dopo una richiesta di chiarimento da parte del bibliotecario, viene spiegato con la ricerca de "la pagina di login per studenti". Ma quando viene illustrato il percorso di login in MyUnito, l'utente segnala con un "qui" la pagina in cui sta agendo, cioè la pagina del sito www.sba.unito.it[16]. Quindi la sua attività è relativa ai servizi bibliografici, in particolare, come si vede poche righe dopo, alla ricerca di un libro.
Questi brevi scambi mostrati, testimoniano quanto sia complicato per gli utenti spiegare cosa stanno cercando di fare in un ambiente web e con un mezzo quale il computer.

[14] Bibliopass ha una funziona analoga a quella del proxy, ma ha una modalità di utilizzo più semplice per l'utente, che inserendo le sole credenziali istituzionali può navigare tra le risorse sottoscritte da UniTO anche se accede da reti esterne. Gli operatori della chat collaborano con gli amministratori di Bibliopass nell'assistenza agli utenti.
[15] Potremmo parlare anche in questo caso di un'estensione di significato: si parla di "catalogo" per intendere uno strumento che permette di cercare tra i documenti offerti dal sistema bibliotecario, indipendentemente dalla tipologia di oggetto che restituisce nella lista dei risultati e dalla modalità di funzionamento
[16] SnapEngage permette di visualizzare, tramite un URL, qual è il widget da cui l'utente sta scrivendo all'agente. In questo caso stava scrivendo dalla chat disponibile sul sito www.sba.unito.it

Ecco che si riafferma la necessità di inferenza da parte del bibliotecario di reference: tra i dipendenti dell'Università è diffusa e nota l'espressione "portale"[17] per indicare il sito www.unito.it. Per gli utenti invece il termine che lo identifica è semplicemente "sito"[18]; formalmente questo non è sbagliato, ma mostra come individui appartenenti allo stesso contesto istituzionale, attribuiscono nomi diversi a ciò con cui lavorano. Il bibliotecario, quindi, deve mostrare elasticità, interpretare la domanda che gli viene posta anche usufruendo di abilità linguistiche. Ciò è naturalmente favorito solo nel momento in cui chi fornisce il servizio di reference ha un quadro generale di quello che l'istituzione in cui è inserito offre e se ha una visione completa dei servizi e delle risorse del sistema bibliotecario nel quale lavora.

Le dinamiche

Qualunque bibliotecario abbia lavorato al bancone ha ben presente la sensazione di timore, incertezza o imbarazzo che molti utenti hanno quando si rivolgono a loro. Ebbene il bancone e la chat box sono entrambi un filtro e il bibliotecario parte sempre da una posizione di vantaggio perché sa per esperienza quello che potrebbe sentirsi chiedere, mentre l'utente spesso non sa con chiarezza ciò che gli serve e soprattutto non sa come formulare la domanda.

È chiaro che il bibliotecario esperto sa che in questi casi è fondamentale ricorrere alle tecniche di intervista[19] per riuscire a comprendere il reale bisogno informativo di chi ha di fronte (o dall'altro capo della rete). Studi come quello di Hirko e Ross[20] e di Radford et al.[21] Dimostrano come sia necessario curare lo scambio tra bibliotecario e utente. La differenza tra una conversazione "in presenza" e una online è che l'empatia, che deve caratterizzare il servizio di reference e su cui Ranganathan tanto fa leva[22], deve trovare in rete uno spazio ancora maggiore.

Il fatto che si stia conversando attraverso una chat box non implica necessariamente una distanza tra gli interlocutori, anche se questa spesso è la percezione che si ha dell'assistenza sincrona digitale[23]. Il tono e gli atteggiamenti professionali del bibliotecario di reference sono gli stessi adottati a un bancone, così come l'utente avrà una certa insicurezza nel chiedere aiuto. L'importante è ricordare

---

[17] "Un portale web, o portale internet (informalmente anche portale) è un sito web che costituisce un punto di partenza, una porta di ingresso, ad un gruppo consistente di risorse in internet o di una intranet" (tratto da Autori di Wikipedia, "Portale web", Wikipedia, L'enciclopedia libera, https://it.wikipedia.org/wiki/Portale_web (controllata il: febbraio 5, 2018)).

[18] "Un sito web o sito Internet (anche abbreviato in sito se chiaro il contesto informatico) è un insieme di pagine web correlate, ovvero una struttura ipertestuale di documenti che risiede su un server web" (tratto da Autori di Wikipedia, "Sito web", Wikipedia, L'enciclopedia libera, https://it.wikipedia.org/wiki/Sito_web (controllata il: febbraio 5, 2018)).

[19] RUSA, «Guidelines for Behavioral Performance of Reference and Information Service Providers», Text, Reference & User Services Association (RUSA), 2013, http://www.ala.org/rusa/resources/guidelines/guidelinesbehavioral.

[20] Buff Hirko e Mary Ross Bucher, *Virtual Reference Training: The Complete Guide to Providing Anytime Anywhere Answers* (Chicago: American Library Association, 2004).

[21] Marie L. Radford et al., «"Are We Getting Warmer?" Query Clarification in Live Chat Virtual Reference», *Reference & User Services Quarterly* 50, n. 3 (2011): 259–79.

[22] Ranganathan, *Il servizio di reference*, 45.

[23] Jody Condit Fagan e Christina M. Desai, «Communication Strategies for Instant Messaging and Chat Reference Services», *The Reference Librarian* 38, n. 79–80 (20 novembre 2002): 121–55, https://doi.org/10.1300/J120v38n79_09.

che il bibliotecario, anche se "digitale", ha una funzione di guida, di accompagnatore oltre che di "segugio" dell'informazione. È vero anche che quest'ultimo aspetto della professione si è fatto ancora più impellente, vista l'enorme mole di informazioni e la varietà di servizi per la ricerca bibliografica disponibili grazie allo sviluppo del web.

Connaway e Faniel hanno realizzato un'importante riattualizzazione delle cinque leggi della biblioteconomia di Ranganathan, considerando la necessità imprescindibile di modernizzare l'approccio alla ricerca bibliografica e ponendo questa volta in cima la regola "Risparmia il tempo del lettore"[24]: questo potenziando le competenze dei bibliotecari nel destreggiarsi con le nuove tecnologie, nelle infrastrutture che le sorreggono e nel mare magnum delle informazioni digitali. Il nuovo fattore chiave nell'assistenza agli utenti è la "*convenience*": il bibliotecario deve inserirsi nel flusso delle attività di ricerca dell'utente, anticiparne i bisogni e selezionare le risorse più adatte ai suoi scopi e facilitare l'accesso a esse.

Ci permettiamo di aggiungere però un altro aspetto che non fa parte dei cinque pilastri del servizio di reference, ma che è comunque uno degli elementi chiave e forse è costitutivo di quell'approccio olistico sostenuto da Ranganathan[25].

Proprio per la necessità di riscattarsi con l'uso, tutti gli oggetti meccanici, che ci circondano possono in ultima analisi essere considerati come oggetti supplichevoli che ci chiamano continuamente in causa. Anzi: per la loro assenza di vita incombono su di noi con un costante senso di morte, se noi stessi non forniamo loro un apporto di energia vitale.

Il reference sincrono è il naturale sviluppo dell'assistenza a un utente che naviga e cerca in rete. Gli strumenti di reference sono per natura artificiosi e la nuova veste che gli è stata data, cioè quella digitale, costituiscono un ulteriore gradino da affrontare per coloro che devono utilizzarli. Capire come aiutare i nostri utenti a salire quel gradino e poter quindi trasmettere loro tutte quelle informazioni che li renderà "information literate" fa parte del compito del digital reference librarian.

Il bibliotecario di reference e il linguista computazionale: le ragioni di una fruttuosa collaborazione

All'interno dell'ambito operativo in cui si articola l'attività di reference bibliotecario in remoto tramite mezzi informatici, emergono con forza quattro grandi elementi disciplinari strettamente correlati e integrati fra loro. Il primo in ordine di importanza fa riferimento ai principi di biblioteconomia teorica e pratica, in quanto elementi indispensabili per la funzione stessa del reference. Poiché l'attività di supporto all'utenza avviene sfruttando sistemi elettronici che operano in remoto, il secondo aspetto interdisciplinare dovrà necessariamente riferirsi al mezzo utilizzato

[24] Lynn Silipigni Connaway e Ixchel M. Faniel, «Reordering Ranganathan: Shifting User Behaviors, Shifting Priorities», 2014, 8, https://www.oclc.org/research/publications/library/2014/oclcresearch-reordering-ranganathan-2014-overview.html.

[25] Ranganathan, *Il servizio di reference*, 66.

per lo scambio comunicativo, ossia, nel caso specifico, la Live Chat sincrona, che si colloca all'interno della vasta area della cosiddetta Comunicazione Mediata dal Computer (CMC). Il terzo elemento che deve essere necessariamente considerato, si collega alla natura propria dei dati prodotti nel corso dell'attività di reference. A tale proposito, è importante sottolineare che lo scambio di informazioni avviene tramite una tecnologia che utilizza unicamente il segno linguistico scritto[26]. Pertanto, ai fini della strutturazione di un processo volto alla migliore comprensione delle dinamiche comunicative tra utente e bibliotecario, tali informazioni dovranno essere sottoposte a tutte le metodologie proprie dell'analisi linguistica che spaziano dalla morfosintassi alla pragmatica. Infine, l'ultimo elemento interdisciplinare che si integra necessariamente con i precedenti, è rappresentato dalle scienze statistiche e dell'informazione. La necessità di elaborare in maniera rapida e automatica l'intera mole di scambi salvati nei file di log del servizio di chat riducendo al minimo la necessità dell'intervento umano, fa sì che questo dovrà limitarsi all'individuazione e alla successiva implementazione di modelli di analisi linguistica attraverso l'uso di software e linguaggi di programmazione dedicati.

L'analisi automatica dei dati della chat

Al fine di realizzare l'analisi dei log della chat in maniera produttiva, è necessario organizzare il lavoro sulla base di una sequenza logica precisa e pianificata. In primo luogo, infatti, è indispensabile avere ben chiaro l'obiettivo finale che si intende raggiungere. Definito questo punto fondamentale, si dovranno quindi individuare gli elementi capaci di soddisfare le esigenze di progetto e che saranno oggetto dell'effettivo lavoro di estrazione. A tale riguardo, infatti, è importante rimarcare come in assenza di solidi criteri di base, il rischio di produrre risultati lacunosi o comunque poco significativi assuma una consistenza tale da minare la validità scientifica dell'indagine stessa. Tuttavia, quanto appena indicato non comporta automaticamente la necessità di dotarsi di fonti informative diverse per rispondere a obiettivi specifici: salvo casi estremamente particolari, è possibile soddisfare una pluralità di indagini collegate a progetti differenti semplicemente modificando i criteri di selezione dei dati appartenenti a una stessa origine. A dimostrazione di questo, possiamo citare due progetti condotti dagli operatori dei Servizi Bibliografici Digitali dell'Università di Torino volti a migliorare la comunicazione con l'utenza attraverso la redazione di FAQ[27] e l'allargamento dell'attività di reference online alla biblioteca di psicologia[28].

Una volta determinati gli obiettivi e le modalità di progetto, il passo successivo consisterà nel valutare quale sistema di estrazione e analisi delle informazioni sia preferibile adottare. Per

[26] Elena Pistolesi, *Il parlar spedito. L'italiano di chat, e-mail e SMS* (Padova: Esedra, 2004), 60.
[27] Muzzupapa, Tomatis, e Bungaro, «Live chat e natural language processing in sinergia per il miglioramento dei servizi bibliotecari.»
[28] Bungaro, Muzzupapa, e Tomatis, «Extending the Live Chat Reference Service at the University of Turin – A Case Study».

soddisfare questo proposito sono possibili due opzioni distinte: la prima consiste nell'individuare sul mercato un prodotto il più possibile adeguato alle esigenze di elaborazione richieste; la seconda prevede la progettazione e l'implementazione di un proprio algoritmo. Per quanto praticabili, entrambe le soluzioni comportano una serie di vantaggi e svantaggi. A tale riguardo, è bene chiarire che anche qualora ci si orienti verso un prodotto commerciale, è condizione imprescindibile individuare la soluzione più consona all'interno della vasta offerta che il mercato dell'informatica propone. In diversi studi quali quelli condotti da Morais & Sampson[29] oppure Fan & Welch[30] si nota l'utilizzo di strumenti semplici quali il foglio di calcolo Excel, mentre altrove[31] si è preferito un database quale Access. Per quanto questi applicativi si rivelino estremamente comodi per condurre indagini relativamente semplici, i progetti appena citati evidenziano la necessità di operare un grosso lavoro manuale di preparazione dei dati nel caso si intenda condurre analisi più raffinate. Un aiuto in proposito giunge dai più complessi software di analisi del contenuto quali Dedoose: questo genere di prodotti consente di concentrarsi sui risultati dell'indagine senza dover prestare particolare attenzione al procedimento che li ha generati[32]. A fronte dei loro evidenti vantaggi, è altresì vero che i sistemi proprietari sono solitamente progettati per coprire una serie predefinita di funzioni possibili e, pertanto, presentano caratteristiche tali da non consentire un rapido adattamento a tutte le esigenze che l'utente può essere intenzionato ad esprimere.

Proprio per ovviare a quest'ultimo aspetto, adottare una soluzione ad hoc rappresenta la risposta più adeguata per quei progetti che richiedono un'elevata versatilità nell'elaborazione dei dati. Questa scelta, per quanto vantaggiosa sotto numerosi aspetti, implica la necessità di disporre di personale dotato sia di competenze tecniche relative alla capacità di progettazione ed implementazione di un determinato algoritmo, sia di tutte le competenze disciplinari necessarie per far sì che il software stesso soddisfi lo scopo per cui è stato creato. Nel caso di indagini che coinvolgono informazioni di natura tipicamente testuale, quali i log della chat di reference, la figura professionale richiesta non potrà corrispondere semplicemente a un esperto in scienze e tecnologie informatiche, ma dovrà comprendere anche tutte le competenze proprie di un linguista. Ecco quindi che laddove il bibliotecario di reference assume il ruolo chiave di esperto capace di fornire adeguata interpretazione tanto delle dinamiche di interscambio con l'utenza, quanto del lessico specialistico utilizzato, la responsabilità dell'adozione di una corretta strategia di estrazione ed analisi delle

[29] Yasmin Morais & Sara Sampson (2010). A content analysis of chat transcripts in the Georgetown Law Library. *Legal Reference Services Quarterly*, 29:3, 165-178.

[30] Suhua Caroline Fan & Jennifer M. Welch (2016) Content Analysis of Virtual Reference Data: Reshaping Library Website Design, *Medical Reference Services Quarterly*, 35:3, 294-304.

[31] Van Houlson, Kate Mccready, e Carla Steinberg Pfahl, «A Window into Our Patron's Needs: Analyzing Data from Chat Transcripts», *Internet Reference Services Quarterly* 11, n. 4 (2007): 19–39, https://doi.org/10.1300/J136v11n04_02.

[32] Michael Mungin, «Stats Don't Tell the Whole Story: Using Qualitative Data Analysis of Chat Reference Transcripts to Assess and Improve Services», *Journal of Library & Information Services in Distance Learning* 11, n. 1–2 (2017): 25–36, https://doi.org/10.1080/1533290X.2016.1223965.

informazioni, nonché lo sviluppo del relativo algoritmo, saranno compiti delegati al linguista computazionale.

L'elaborazione dei dati: aspetti teorici e risvolti pratici

Avendo delineato le possibili alternative inerenti le scelte operative ed i relativi ruoli, verrà ora dedicato spazio alla descrizione delle strategie utili per l'analisi dei testi (Text Mining) attraverso l'elaborazione automatica del linguaggio naturale (NLP – Natural Language Processing[33]). A tale proposito è bene chiarire lo stretto rapporto esistente tra i due approcci. Le strategie di NLP consentono di desumere le proprietà del testo sotto il profilo linguistico: questi dati risultano funzionali alle operazioni di Text Mining che, per contro, sono principalmente orientate all'estrazione di nuova informazione anche in termini statistici. Sulla base di tali premesse, quindi, sarà necessario definire una sequenza di elaborazioni del testo oggetto di analisi secondo un preciso schema a *step* in cui ogni passo opererà sul risultato di tutte le operazioni precedentemente svolte. Dato il carattere propedeutico di ogni singola elaborazione, l'intera filiera prevede che nelle fasi iniziali vengano poste in essere tutte le operazioni orientate alla preparazione del testo. In termini pratici, questo comporta l'eventuale normalizzazione degli elementi di cui il testo è composto, che si traduce nell'attività di riallineamento o ricomposizione di strutture quali righe o paragrafi che, per varie ragioni, possono presentare una distribuzione non lineare ed omogenea. Sempre all'interno della preparazione del testo, ricopre aspetti di interesse e utilità decisamente maggiori l'azione di *tokenizzazione*[34]. Tale attività, che rappresenta un passaggio imprescindibile per poter svolgere qualsivoglia operazione di trattamento automatico del linguaggio, consiste nell'isolare con chiarezza ogni singola unità lessicale di cui il testo è composto. L'obiettivo in questione, tuttavia, potrà essere raggiunto unicamente attraverso il superamento delle regole ortografiche comunemente utilizzate in una determinata lingua. Per quanto apparentemente di semplice realizzazione, la progettazione di un sistema di tokenizzazione può rivelarsi, in realtà, particolarmente problematico. L'esistenza di regole ortografiche distinte per ogni lingua sancisce, infatti, l'impossibilità di realizzare un algoritmo universale capace di adattarsi a qualsiasi testo. In qualità di esempio, l'utilizzo, seppur sporadico, dell'inglese nella sua funzione di lingua franca[35] all'interno del servizio di reference tramite Live Chat, rende complessivamente più difficoltosa la suddivisione in token dei messaggi. Per quanto, da un punto di vista tipologico, le caratteristiche dell'inglese siano tali da poterlo addirittura ricondurre a una lingua isolante come il cinese[36], che non richiede alcun tipo di elaborazione preparatoria del testo, per alcuni aspetti puntuali quali le forme contratte "isn't" o

[33] Anne Kao e Steve R. Poteet, *Natural Language Processing and Text Mining* (London: Springer London, 2007).
[34] Alessandro Lenci, Vito Pirrelli, e Simonetta Montemagni, *Testo e computer : elementi di linguistica computazionale* (Roma: Carocci, 2016), 102–22.
[35] Barbara Seidlhofer, *Understanding English as a Lingua Franca* (Oxford: Oxford University Press, 2011).
[36] Antonio Romano e Anna Maria Miletto, *Argomenti scelti di glottologia e linguistica* (Torino: Omega, 2010), 142.

"won't", risulta inevitabile integrare il sistema di regole capace di gestire le norme ortografiche dell'italiano con quelle utili per manipolare questo genere di particolarità.
Un ulteriore livello di criticità è dato dalle variazioni diafasiche[37] che caratterizzano l'interazione fra parlanti. Questo parametro può risultare particolarmente significativo nel momento in cui l'interscambio comunicativo registra l'utilizzo dei cosiddetti emoticon[38], detti anche "faccine" in italiano, elementi testuali dotati di un grado di iconicità molto elevato. Appartenenti alla funzione emotiva dal punto di vista della classificazione di Jackobson[39], il loro uso viene fatto risalire alle prime forme di posta elettronica realizzate all'inizio degli anni '80 del XX secolo al fine di sopperire all'assenza di quegli elementi extralinguistici che caratterizzano l'interazione faccia a faccia quali gestualità e mimica del volto. Inizialmente gli emoticon potevano essere rappresentati unicamente mediante una stringa di caratteri tipografici solo in teoria rigidamente codificata, ma di fatto soggetta a numerose possibili modifiche ed estensioni. Proprio a causa di ciò, le faccine rappresentano costrutti particolarmente ardui da individuare e isolare mediante sistemi automatici in quanto possono essere facilmente confusi con regolari elementi di punteggiatura. In qualità di esempio, è possibile notare come all'interno della seguente riga sia difficile distinguere l'emoticon ":))" dalla corretta chiusura della parentesi:

Diego: 11:06:08: a dopo (se necessario! :))

Relativamente a quanto appena indicato, è comunque bene ricordare che nel corso degli anni, a seguito della spinta tecnologica orientale legata alla telefonia cellulare, gli emoticon hanno subito importanti evoluzioni, trasformandosi da semplici formule testuali in oggetti grafici di varia complessità denominati "emoji", alcuni dei quali addirittura dotati di caratteristiche dinamiche.
Entrando maggiormente nel dettaglio, descriviamo ora la tipologia di emoticon presenti all'interno degli interscambi di reference bibliotecario mediante Live Chat. In prima istanza, è possibile suddividere queste particolari espressioni in tre gruppi distinti in relazione alle loro caratteristiche. Sulla base delle 2056 occorrenze registrate nel corpus delle chat, il primo raggruppamento in ordine di importanza raccoglie al suo interno tutte le faccine orientate sul piano orizzontale tra cui quelle sorridenti :-) :) ammiccanti ;-) ;) oppure tristi :(. Il secondo gruppo, per contro, è rappresentato dai cosiddetti "kaomoji", faccine caratterizzate dal tipico orientamento verticale quale ^__^ che rappresenta felicità oppure >.< indice di fastidio. Originati in seno al contesto culturale nipponico, i kaomoji rappresentano una realtà molto complessa anche dal punto di vista realizzativo, in quanto

[37] Mirko Tavosanis, *L'italiano del web* (Roma: Carocci, 2011), 33–36.
[38] Stefania Spina, «Emoticons as Multifunctional and Pragmatic Resources: A Corpus-Based Study on Twitter» (cmc-corpora conference series, 2017), https://doi.org/10.5281/zenodo.1041883.
[39] Roman Jakobson, «Linguistics and poetics», in *Style in language* (New York; London: Technology Press of MIT : John Wiley & Sons, 1960), 350–77.

richiedono l'utilizzo di alcuni grafemi[40] tipici della lingua giapponese. Proprio per tale ragione la loro diffusione risulta decisamente ridotta: all'interno del corpus, infatti, si sono potute rilevare appena 10 occorrenze, tutte caratterizzate da una struttura estremamente semplice. Un aspetto interessante che accomuna i due gruppi appena descritti è dato dalla perdita completa sia del valore ortografico veicolato dalle interpunzioni, sia del valore fonologico rappresentato dal grafema: all'interno della struttura iconica dell'emoticon i caratteri trascendono la loro funzione originale, assumendo una dimensione esclusivamente grafica[41]. Il terzo ed ultimo gruppo, infine, è costituito dagli emoji grafici resi disponibili all'utenza dal programma di chat a partire dal 2016. Queste faccine, trascritte automaticamente in formato testuale dal software di gestione durante la produzione del file di log, registrano un totale di 98 occorrenze.

Poiché un'analisi dettagliata in termini quantitativi e qualitativi su tutti gli emoticon presenti nel corpus richiederebbe uno spazio dedicato, ci limiteremo a esaminare quello utilizzato con maggiore frequenza. Nato come stilizzazione della faccina che ride ("smile" in inglese), l'emoticon in questione presenta due varianti: :-) e :) . Verosimilmente in virtù del fatto che la sua digitazione richiede minor tempo, la seconda espressione risulta quella più diffusa, con 1590 occorrenze rispetto alle sole 116 della variante originale. Oltre a ciò, è interessante rilevare che analogamente ad alcuni elementi della punteggiatura quali il punto esclamativo e interrogativo, la cui valenza espressiva può essere modulata nella scrittura informale attraverso il meccanismo della ripetizione, anche gli emoticon possono esprimere gradi diversi di intensità mediante la moltiplicazione del carattere terminale. In qualità di esempio, si noti come l'emoticon " :((((" venga utilizzato per trasmettere il senso di estrema insoddisfazione e disappunto:

Maria Vittoria : 14:07:18: è probabile che ci sia molto traffico in questo momento
Maria Vittoria : 14:07:30: prova ad aspettare un pochino
Visitor : 14:07:50: eh lo so ...
Visitor : 14:08:01: però ha smesso di funzionare a mezzogirno
Visitor : 14:08:14: e io devo consegnare il lavoro domani !! :((((

Seppur con funzioni differenti, è possibile ritrovare l'emoticon del sorriso tanto tra gli utenti, quanto tra gli operatori. Infatti, laddove il bibliotecario di reference utilizza questa faccina fondamentalmente per mettere a proprio agio l'interlocutore spostando l'interscambio su un piano meno formale e comunicargli la propria disponibilità all'ascolto, per l'utente, invece, oltre a confermare l'approccio informale del dialogo, rappresenta un modo per esprimere sia il proprio imbarazzo di fronte a qualche difficoltà tecnica, sia la soddisfazione per la risoluzione del problema per cui ha chiesto sostegno. Quanto appena descritto dimostra che, unitamente ad altri parametri linguistici che esamineremo successivamente, l'analisi dell'utilizzo degli emoticon costituisce una

[40] M. Luisa Porzio Gernia e Valeria Lomanto, *Foni, fonemi, grafemi* (Torino: Giappichelli, 1983), 121–24.
[41] Il carattere maiuscolo "D" costitutivo dell'emoticon ":-D " intende rappresentare una risata a bocca aperta e non l'occlusiva dentale sonora /d/.

fonte tangibile per condurre in maniera indiretta la misura del grado di soddisfazione del servizio di reference da parte dell'utenza.

<chat>
Visitor : 14:59:20: ciao , si può considerare TUTTO una piattaforma online ? se è corretto il termine :)
Maria Vittoria : 15:01:08: no
Visitor : 15:01:31: grazie
Maria Vittoria : 15:01:31: non esattamente
Maria Vittoria : 15:01:44: è un enorme database
Maria Vittoria : 15:02:02: che indicizza record bibliografici
Maria Vittoria : 15:02:22: per permettere una ricerca come se fosse un motore di ricerca
Visitor : 15:02:41: perfetto grazie mille !!! :)
Maria Vittoria : 15:02:56: :)
</chat>

Oltre alle funzioni fin qui indicate, l'algoritmo di tokenizzazione può anche includere l'individuazione di strutture lessicali complesse quali nomi propri, titoli, unità polirematiche e, in generale, tutte quelle forme il cui significato è espresso da un significante costituito da una sequenza precisa di lessemi distinti. Ovviamente la scelta di quali espressioni multilessicali debbano essere prese in considerazione è del tutto discrezionale e deve essere svolta in funzione del risultato che si intende raggiungere. Se da un lato è opportuno che i marchi commerciali (es. "Coca Cola"), gli elementi di toponomastica (es. "La Spezia") e i nomi propri di persone o istituzioni quali "Ettore Fico" o "Mole Antonelliana" vengano riconosciuti e trattati in maniera analoga a sostantivi quali "banca dati" o "motore di ricerca", è possibile considerare i titoli di libri o riviste e i messaggi di errore o di sistema come costituiti da lessemi indipendenti oppure gestirli come forme unitarie, qualora lo si ritenga più utile. In qualità di esempio, nel caso dei progetti di analisi automatica degli scambi relativi al servizio di chat operato dal Sistema Bibliotecario dell'Università di Torino, la scelta è ricaduta su una modalità operativa che tratta come espressioni multilessicali univoche tutte quelle parti di testo che non rientrano nella produzione originale degli interlocutori. Tale soluzione costituisce il frutto di un scelta mirata orientata a far sì che i dati statistici prodotti dalle operazioni di text mining siano condizionati il meno possibile da fattori esterni capaci di influenzare l'indagine sul linguaggio effettivamente utilizzato da utenti e operatori del reference online. Al fine di automatizzare quanto più possibile l'individuazione di titoli e messaggi di sistema, è possibile seguire un approccio basato sull'analisi delle concordanze: tale soluzione consente di far emergere la presenza di co-occorrenze lessicali stabili, definite in linguistica con il termine "collocazione"[42]. Esaminando il contesto frasale, è quindi possibile riscontrare come certi elementi siano solitamente accompagnati da termini specifici quali "titolo" o "messaggio", che per la loro rilevanza assumono quindi il ruolo di parole chiave.

[42] Thierry Fontenelle, «What on earth are collocations?», *English Today* 10, n. 4 (1994): 42–48, https://doi.org/10.1017/S0266078400007902.

Es:

Maria Vittoria : 09:51:38: mi dai il titolo dell' articolo ?
Visitor : 09:51:59: fear of crime in residential communities
Visitor : 09:14:46: ho trovato il mio articolo su trova unito ma ...
Visitor : 09:15:05: mi dice questo messaggio :
Visitor : 09:15:09:
ATTENZIONE: Consultabile esclusivamente dalle postazioni della Scuola di Medicina o attraverso il proxy di Medicin a .

Una volta terminata la preparazione del testo avvalendosi delle metodologie poc'anzi descritte, è possibile effettuare l'analisi vera e propria dei testi operando una valutazione statistica delle componenti lessicali. A tale fine, è utile fare riferimento a metodologie consolidate quali "la legge di Zipf"[43], che consente di ottenere un profilo chiaro in termini quantitativi e qualitativi delle caratteristiche del testo esaminato. Poiché secondo Zipf vi è una precisa relazione tra la frequenza di un dato termine e il suo rango, ossia la posizione ricoperta all'interno della lista di tutte le parole utilizzate nel testo, sarà necessario dotarsi di un algoritmo adatto allo scopo. Oltre a questo, un ulteriore elemento indicativo delle caratteristiche generali del testo è fornito dal rapporto type/token[44]. Il valore in questione, che deve essere compreso tra 0 e 1, indica il grado di variabilità e ricchezza del lessico: tanto più il valore sarà prossimo all'unità, tanto maggiore sarà l'uso di termini poco ricorrenti. Nel caso dei log di reference esaminati nel loro complesso i type corrispondono a 28203 elementi, mentre i token a 474200 occorrenze. Il loro rapporto risulta corrispondere a 0,06 e dimostra che il testo analizzato è lessicalmente piuttosto povero e si caratterizza per la presenza di numerose forme ripetute. Al fine di comprendere meglio la natura del dato in questione, si rivela utile analizzare separatamente i valori di type e token computati sulle frasi prodotte dagli utenti e dagli operatori della chat. A seguito di tale calcolo è possibile notare come le produzioni di tutti gli utenti registrino la presenza di 15233 type e 173411 token, il cui rapporto risulta pari a 0,09. Per contro, i valori riconducibili agli enunciati dei soli operatori di reference consistano rispettivamente in 12970 type e 300789 token, con un rapporto di 0,04. Proprio il confronto dei dati appena descritti dimostra chiaramente che la responsabilità della scarsa variabilità lessicale ricade sugli operatori, la cui produzione è caratterizzata da informazioni di natura prevalentemente tecnica e, pertanto, più stabile e standardizzata.

Per quanto concerne l'aspetto qualitativo, è utile ricordare che, sempre secondo Zipf, la sua valutazione è funzione del dato quantitativo. Scremando, infatti la punteggiatura e le parole appartenenti alle classi chiuse quali articoli, congiunzioni, pronomi e avverbi caratterizzati da una frequenza d'uso molto elevata, è possibile notare alcune interessanti discrepanze in termini di occorrenze tra il lessico impiegato dagli utenti rispetto a quello degli operatori. A titolo di esempio, è curioso osservare che la parola "grazie" è utilizzata ben 5464 volte dagli utenti, a fronte di sole

[43] Steven Piantadosi, «Zipf's Word Frequency Law in Natural Language: A Critical Review and Future Directions», *Psychonomic Bulletin & Review* 21, n. 5 (2014): 1112–1130, https://doi.org/10.3758/s13423-014-0585-6.
[44] Lenci, Pirrelli, e Montemagni, *Testo e computer : elementi di linguistica computazionale*, 133.

1397 occorrenze del “prego” di risposta dell’operatore. Un tale uso massiccio del termine di gratitudine, spesso legato all’elemento rafforzativo “mille”, che registra 1522 occorrenze, segnala la fiducia dell’utente nei confronti dell’operatore e conferma quell’elevato livello di soddisfazione del servizio già riscontrato dall’analisi degli emoticon.

References

Anne Kao, e Steve R. Poteet. *Natural Language Processing and Text Mining*. London: Springer London, 2007.

Bungaro, Franco, Maria Muzzupapa, e Marco Tomatis. «Extending the Live Chat Reference Service at the University of Turin – A Case Study». *Proceedings of the IATUL Conferences*, 2017. https://docs.lib.purdue.edu/iatul/2017/challenges/1.

Connaway, Lynn Silipigni, e Ixchel M. Faniel. «Reordering Ranganathan: Shifting User Behaviors, Shifting Priorities», 2014. https://www.oclc.org/research/publications/library/2014/oclcresearch-reordering-ranganathan-2014-overview.html.

Fagan, Jody Condit, e Christina M. Desai. «Communication Strategies for Instant Messaging and Chat Reference Services». *The Reference Librarian* 38, n. 79–80 (20 novembre 2002): 121–55. https://doi.org/10.1300/J120v38n79_09.

Fontenelle, Thierry. «What on earth are collocations?» *English Today* 10, n. 4 (1994): 42–48. https://doi.org/10.1017/S0266078400007902.

Galeffi, Agnese. «Se il catalogo parlasse, lo capiremmo? Cinque assiomi della comunicazione catalografica». *AIB studi* 57, n. 2 (2017). https://doi.org/10.2426/aibstudi-11648.

Hirko, Buff, e Mary Ross Bucher. *Virtual Reference Training: The Complete Guide to Providing Anytime Anywhere Answers*. Chicago: American Library Association, 2004.

Houlson, Van, Kate Mccready, e Carla Steinberg Pfahl. «A Window into Our Patron’s Needs: Analyzing Data from Chat Transcripts». *Internet Reference Services Quarterly* 11, n. 4 (2007): 19–39. https://doi.org/10.1300/J136v11n04_02.

IFLA. «IFLA Digital Reference Guidelines», 2002. https://www.ifla.org/files/assets/reference-and-information-services/publications/ifla-digital-reference-guidelines-en.pdf.

Jakobson, Roman. «Linguistics and poetics». In *Style in language*, 350–77. New York; London: Technology Press of MIT : John Wiley & Sons, 1960.

Lankes, R. David. «New Concepts in Digital Reference». *Synthesis Lectures on Information Concepts, Retrieval, and Services* 1, n. 1 (2009). http://www.morganclaypool.com/doi/abs/10.2200/S00166ED1V01Y200812ICR001.

Lenci, Alessandro, Vito Pirrelli, e Simonetta Montemagni. *Testo e computer : elementi di linguistica computazionale*. Roma: Carocci, 2016.

Luo, Lili. «Chat reference competencies: identification from a literature review and librarian interviews». *Reference Services Review* 35, n. 2 (2007): 195–209. https://doi.org/10.1108/00907320710749137.

Mungin, Michael. «Stats Don’t Tell the Whole Story: Using Qualitative Data Analysis of Chat Reference Transcripts to Assess and Improve Services». *Journal of Library & Information Services in Distance Learning* 11, n. 1–2 (2017): 25–36. https://doi.org/10.1080/1533290X.2016.1223965.

Muzzupapa, Maria Vittoria, Marco Stefano Tomatis, e Franco Carlo Bungaro. «Live chat e natural language processing in sinergia per il miglioramento dei servizi bibliotecari.» *Biblioteche oggi Trends* 3, n. 1 (2017): 41–53. https://doi.org/10.3302/2421-3810-201701-041-1.

Piantadosi, Steven. «Zipf’s Word Frequency Law in Natural Language: A Critical Review and Future Directions». *Psychonomic Bulletin & Review* 21, n. 5 (2014): 1112–1130. https://doi.org/10.3758/s13423-014-0585-6.

Pistolesi, Elena. *Il parlar spedito. L'italiano di chat, e-mail e SMS*. Padova: Esedra, 2004.
Porzio Gernia, M. Luisa, e Valeria Lomanto. *Foni, fonemi, grafemi*. Torino: Giappichelli, 1983.
Radford, Marie L., Lynn Silipigni Connaway, Patrick A. Confer, Susanna Sabolcsi-Boros, e Hannah Kwon. «"Are We Getting Warmer?" Query Clarification in Live Chat Virtual Reference». *Reference & User Services Quarterly* 50, n. 3 (2011): 259–79.
Ranganathan, S. R. *Il servizio di reference*. Tradotto da Carlo Bianchini. Firenze: Le Lettere, 2010.
Romano, Antonio, e Anna Maria Miletto. *Argomenti scelti di glottologia e linguistica*. Torino: Omega, 2010.
RUSA. «Guidelines for Behavioral Performance of Reference and Information Service Providers». Text. Reference & User Services Association (RUSA), 2013. http://www.ala.org/rusa/resources/guidelines/guidelinesbehavioral.
———. «Guidelines for Implementing and Maintaining Virtual Reference Services». Text. Reference & User Services Association (RUSA), 2017. http://www.ala.org/rusa/resources/guidelines/virtrefguidelines.
Seidlhofer, Barbara. *Understanding English as a Lingua Franca*. Oxford: Oxford University Press, 2011.
Spina, Stefania. «Emoticons as Multifunctional and Pragmatic Resources: A Corpus-Based Study on Twitter». cmc-corpora conference series, 2017. https://doi.org/10.5281/zenodo.1041883.
Tavosanis, Mirko. *L'italiano del web*. Roma: Carocci, 2011.