

Tesi di Laurea in Ingegneria Informatica
Indirizzo "Automatica e sistemi di automazione industriale"

Applicazione di tecniche e strumenti di comunicazione multimodale e di valutazione della leggibilità e comprensibilità di contenuti web

Politecnico di Milano
V Facoltà di Ingegneria – Campus Leonardo
Anno Accademico 2003-2004



Keywords: accessibilità, portali web, disabilità cognitive, linguistica computazionale, analisi
sintattica di testi, leggibilità e complessità di testi

Relatrice: Licia Sbattella

Correlatori: Ing. Antonio Bianchi e Ing. Giuliano Pozza

Autore: Davide Ceriani

Matricola: 641805

E-mail: davideceriani1@tin.it – dc641805@stud.polimi.it

"I limiti del mio linguaggio significano i limiti del mio mondo"

L. Wittgenstein

Alla mia famiglia: Pierino, Luisa, Elisa e a mio zio Ezio

Sommario

1.	Introduzione.....	1-5
1.1.	Accessibilità	1-5
1.2.	Accessibilità: situazione in Italia	1-8
1.3.	Legge Stanca, una chiave di lettura	1-9
1.3.1.	Limiti della legge Stanca	1-11
1.4.	Limiti ed evoluzioni degli standard attuali	1-12
1.4.1.	Il rapporto WWAAC	1-13
1.5.	Applicazioni degli standard	1-14
1.6.	Impatto dell'accessibilità sui costi e sulle modalità di progettazione delle architetture web: project management per l'accessibilità.....	1-18
2.	Introduzione al progetto "navigAbile©"	2-22
2.1.	La Soluzione proposta.....	2-26
2.2.	L'architettura dei Servizi di navigAbile: introduzione.....	2-27
2.2.1.	Servizi di supporto alla gestione dei contenuti	2-28
2.2.2.	Servizi di Personalizzazione e Navigazione	2-29
2.2.3.	Servizio di Registrazione e profilazione.....	2-30
2.2.4.	La prima applicazione e la validazione scientifica	2-31
2.3.	Architettura dei servizi di navigAbile	2-31
2.4.	Servizi per la Comunicazione Aumentativa Alternativa	2-35
2.4.1.	Semplificazione struttura e contenuti	2-36
2.4.2.	Registrazione.....	2-37
2.4.3.	Personalizzazione	2-38
2.4.4.	Autenticazione	2-40
2.4.5.	Dizionario visuale	2-41
2.4.6.	La Tabella di Comunicazione	2-41
2.4.7.	Compositore di frasi	2-43
2.4.8.	Bacheca.....	2-43
2.4.9.	Forum	2-44
2.4.10.	E-mail	2-44
2.5.	Il prototipo statico iniziale	2-44
2.6.	Il sito pilota.....	2-49
2.7.	Gestione del progetto: le specificità di navigAbile.....	2-56
2.8.	Sviluppi futuri di navigAbile	2-58
2.9.	Attività a cui ho partecipato e componenti architetture che ho sviluppato	2-61
3.	Il contesto	3-63
3.1.	La Comunicazione Aumentativa Alternativa (CAA).....	3-63
3.2.	Linguaggi iconici: il caso di PCS e BLISS.....	3-64

3.3.	Leggibilità	3-67
3.4.	Gli indici di leggibilità: l'indice di Gulpease	3-69
3.5.	Plain Language: leggibilità e comprensibilità	3-72
3.6.	Vocabolario fondamentale di De Mauro.....	3-73
3.6.1.	L'indicatore di comprensibilità di navigAbile	3-75
4.	Il dizionario Multimodale di navigAbile.....	4-76
4.1.	Struttura e schema concettuale.....	4-76
4.2.	Scenari di utilizzo tipici del dizionario di navigAbile.....	4-82
5.	Analisi di testi.....	5-84
5.1.	Introduzione alla comprensione del linguaggio.....	5-84
5.2.	L'analisi lessicale.....	5-86
5.3.	L'analisi sintattica	5-88
5.4.	L'analisi semantica	5-90
5.5.	I diversi tipi di ambiguità.....	5-92
5.6.	Utilizzo dell'analisi sintattica nel contesto del progetto navigAbile.....	5-94
5.7.	Parser chunk-IT	5-97
5.8.	Una metodologia per la valutazione della leggibilità di un testo	5-99
5.9.	Risultati dell'analisi di leggibilità dei testi di prova	5-104
5.10.	Analisi della traduzione dall'italiano verso i linguaggi iconici.....	5-106
6.	Conclusioni e Future Works	6-109
7.	Bibliografia.....	7-112
7.1.	Riferimenti per il tema dell'accessibilità	7-112
7.2.	Riferimenti di linguistica	7-113
8.	Elenco figure e tabelle	8-114
8.1.	Indice delle figure	8-114
8.2.	Indice delle tabelle	8-115
9.	Ringraziamenti	9-116

1. Introduzione

1.1. Accessibilità

Con accessibilità si intende la facilità con cui un soggetto interessato e autorizzato a conoscere il dato riesce ad accedervi con le tecnologie disponibili, tenuto conto della sua situazione fisica, psichica e culturale.

Accessibilità vuol dire soprattutto e-partecipation, cioè la rimozione delle barriere informatiche. Con questo termine si indicano le difficoltà che i disabili incontrano nell'usare un sistema informatico, che discendono da:

- ✓ Carenze nella progettazione del software e dei contenuti dei siti web, che non tengono conto dei principi della progettazione universale e dell'usabilità, peraltro ormai approfonditamente studiati e oggetto di raccomandazioni e standard a livello internazionale;
- ✓ Indisponibilità dei dispositivi specificamente realizzati per favorire l'uso dei computer da parte dei disabili (per esempio le tastiere e i mouse utilizzati in alcune disabilità motorie, lo screen reader, la barra Braille o i sintetizzatori vocali usati dai non vedenti).

Il primo problema, pur non riguardando soltanto i disabili e la cui soluzione spesso non richiede alcun investimento aggiuntivo, ha notoriamente le conseguenze più negative sull'accessibilità. L'importanza dell'accesso dei disabili alle tecnologie informatiche è stata sottolineata da tempo in numerosi documenti internazionali ed è argomento prioritario nei programmi della Commissione europea.

Un sito **accessibile** è quindi un ambiente multimediale in cui il contenuto informativo, la modalità di interazione, le procedure di navigazione e i servizi disponibili siano pienamente fruibili da qualsiasi tipo di utente indipendentemente dalle sue capacità, dalla sua abilità, dalla configurazione hardware e software disponibile.

I siti devono essere usabili. L'usabilità del web implica che le informazioni debbano essere organizzate e strutturate in maniera da garantire la massima fruibilità.

Un aspetto spesso trascurato è il seguente: il concetto di usabilità si pone ad un livello superiore rispetto a quello di accessibilità. Non è quindi sufficiente che un sito web sia accessibile, cioè che offra la possibilità di un'interazione, ma deve essere garantito che tale interazione sia efficiente.

Il concetto di usabilità deve essere collocato in una prospettiva sistemica: l'usabilità è una caratteristica che dipende da chi lo usa, dal compito da svolgere, dall'interazione con altre attrezzature e dall'ambiente in cui il prodotto è utilizzato.

A livello internazionale, l'ente più rappresentativo preposto alla definizione delle linee guida per il tema dell'accessibilità dei documenti web è il WAI (Web Accessibility Initiative) proposto dal W3C (World Wide Web Consortium). L'iniziativa per l'accesso al web del W3C riunisce persone provenienti da aziende, associazioni sull'handicap, enti statali e laboratori di ricerca per studiare protocolli e programmi che rendano il web accessibile ai portatori di handicap visivi, uditivi, motori e cognitivi; il suo compito spazia dalla revisione delle tecnologie per l'accessibilità sino alla stesura di linee guida per l'accesso a siti web, oltre allo sviluppo di altri strumenti in grado di valutare l'accessibilità (i cosiddetti validatori); nelle linee guida si invita non tanto a limitare l'uso di supporti multimediali, bensì a prevedere che i fruitori del web possano accedere ad un sito con modalità diverse. Alcune di queste norme prevedono, ad esempio, che ogni immagine sia corredata da un equivalente testuale che ne descriva il contenuto, in modo da poter facilitare la navigazione di un utente non vedente che utilizza ausili dedicati (screen reader, display Braille). Un sito accessibile dovrebbe sfruttare i cosiddetti "equivalenti testuali", associati a tutte le componenti non testuali, come grafici, immagini e animazioni, in modo da inserire la ridondanza di informazione che permetta la presentazione alternativa. La versione "solo testo" viene considerata una soluzione di rifugio, quando non sia praticabile quella canonica, e comunque da non privilegiare. Si dovrebbero evitare poi i cambi di colore improvvisi, gli scatti, il refresh automatico, che in alcuni soggetti, come le persone che soffrono di epilessia, possono provocare reazioni negative.

Il principio basilare che sottende l'indipendenza degli strumenti e l'accessibilità è la separazione della forma dal contenuto; quando il significato di un documento è salvato separatamente da come deve apparire, l'indipendenza degli strumenti e l'accessibilità sono più facili da tutelare e questo lo si può ottenere con i fogli di stile (CSS), cioè una serie di istruzioni su come presentare o trasformare una pagina web: ad esempio si definisce la spaziatura e la dimensione del carattere, i colori, ecc.

Esiste anche un'altra serie di norme, non redatte dal WAI ed appartenenti ad una legge federale statunitense, chiamate Section508 Federal Accessibility Guidelines: si tratta di un elenco di regole da rispettare per ottenere la relativa certificazione di accessibilità.

Per certificare il livello di accessibilità e la conformità ad un certo insieme di norme, sono a disposizione degli sviluppatori una serie di strumenti software atti alla verifica di conformità. Alcuni di questi software sono anche in grado di fornire i suggerimenti per ottenere il livello di accessibilità desiderato. Alcuni validatori si integrano in ambienti di sviluppo esistenti per pagine web (ad esempio LIFT for DreamWeaver è un validatore che si integra con il software della MacroMedia). Alcuni strumenti di sviluppo (come GoLive della Adobe) integrano già degli strumenti per la verifica del livello di accessibilità

Tra i validatori più diffusi si possono ricordare:

- Bobby (<http://bobby.watchfire.com/bobby/>): è un servizio gratuito per verificare la conformità di pagine web con le linee guida per l'accessibilità. Disponibile anche una versione commerciale con maggiori funzionalità, ad esempio quelle di riparazione.
- A-Prompt: si tratta di un tool gratuito che consente sia la validazione che la riparazione di pagine HTML.
- LIFT (<http://usablenet.com>): sono diversi strumenti per la validazione e la riparazione, che si integrano nei sistemi di sviluppo più diffusi per documenti web.

Di seguito viene proposta una serie di motivazioni che giustificano opportunità e sostenibilità economica di un progetto web accessibile (Landoni, [1]):

1. La prima motivazione per la realizzazione di un sito web accessibile è certamente il valore etico e sociale del progetto. Questa prerogativa è irrinunciabile per alcune categorie di enti, come quelli statali, che offrono un servizio pubblico e che quindi teoricamente dovrebbe poter essere usufruito da tutti.
2. Un sito accessibile può portare ad un incremento del fatturato, grazie alla possibilità di raggiungere un maggior bacino di utenza.
3. L'attenzione alla tematica sociale dell'accessibilità può costituire un vantaggio competitivo in termini di immagine. Si pensi ad esempio ai colossi dell'e-commerce o ai produttori di sistemi operativi (come ad esempio Amazon e Microsoft).
4. Il disaccoppiamento contenuti-presentazione dei contenuti può aprire la strada alla portabilità del proprio sito, oltre che su browser speciali per disabili, su dispositivi innovativi come WAP, UMTS, ecc., con un conseguente significativo vantaggio in

termini di differenziazione dell'offerta, di utenti raggiungibili e di ritorno di immagine per l'implementazione di soluzioni all'avanguardia.

1.2. Accessibilità: situazione in Italia

In Italia la Pubblica Amministrazione, istituzionalmente più attenta alle problematiche delle minoranze, è tenuta nella sua comunicazione pubblica al cittadino attraverso Internet a rispettare, oltre alle Leggi che tutelano i diritti dei disabili, alcune precise direttive; nel marzo 2001, infatti, l'allora Ministro della Funzione Pubblica Bassanini promulgò una Direttiva (Linee Guida per l'organizzazione, l'usabilità e l'accessibilità dei siti web delle pubbliche amministrazioni, Circolare 13 marzo 2001, n.3/2001) in cui chiedeva il rispetto di alcuni principi che consentissero la piena accessibilità di tutti i siti della Pubblica Amministrazione; nel maggio dello stesso anno, l'AIPA (Autorità per l'Informatica nella Pubblica Amministrazione) costituiva un Gruppo di lavoro sull'Accessibilità e tecnologie informatiche nella Pubblica Amministrazione che ha rilasciato di recente il documento "Criteri e strumenti per migliorare l'accessibilità dei siti web e delle applicazioni informatiche a persone disabili" (Circolare AIPA del 6 settembre 2001, n. AIPA/CR/32) a cui tutte le Istituzioni devono fare riferimento.

Alcuni dati possono dare un'idea della portata del problema, attraverso la stima del numero di persone interessate a questa tematica: secondo una ricerca condotta dall'ISTAT nel 1999 (negli anni successivi l'ISTAT ha modificato le categorie di classificazione dei disabili, ma i dati del '99 sono comunque significativi ai nostri scopi) si può stimare che in Italia ci siano circa 2,9 milioni di persone disabili. La seguente tabella mette in evidenza la ripartizione nelle diverse tipologie di disabilità:

Tipo di disabilità	Numero di disabili in Italia	Percentuale
Motoria	1.000.000 di cui 60.000 in carrozzina	34%
Visiva	350.000	12%
Uditiva	800.000 di cui 50.000 sordomuti	28%
Mentale	750.000	26%

Tabella 1: Distribuzione dei diversi tipi di disabilità

1.3. Legge Stanca, una chiave di lettura

Il 16 ottobre 2003 la Camera dei Deputati ha approvato all'unanimità il testo unificato dei nove progetti di legge presentati in Parlamento a partire dal 2001; entrambi i rami del Parlamento hanno appoggiato senza riserve i principi e le iniziative proposte affinché servizi e opportunità di lavoro offerti dalle nuove tecnologie siano pienamente disponibili per tutti i cittadini.

L'obiettivo di una reale accessibilità riguarda, infatti, non solo i disabili intesi nella accezione tradizionale, ma tutte le categorie deboli ed in particolare quella degli anziani.

Il testo approvato stabilisce l'obbligo per tutte le pubbliche amministrazioni e per alcuni enti para-statali, di conformare il proprio sito Web a criteri di accessibilità. Tale obbligo sussiste in caso sia di "rinnovo, modifica o novazione" del sito Web e dovrà essere concluso entro 12 mesi dall'approvazione del decreto (la legge è stata poi approvata nel Gennaio 2004).

Come detto il testo approvato è unificato e riunisce tutti i nove disegni di legge presentati a partire dal 2001 da altrettanti parlamentari. Tra questi DDL vi è anche quello del ministro per l'Innovazione e Tecnologie, Lucio Stanca, che ha costituito il nucleo centrale del testo attorno cui la commissione ha lavorato per unificare i diversi progetti di legge.

Il testo della legge contiene alcune direttive atte a segnalare i punti essenziali sull'accessibilità: l'organizzazione delle pagine web, immagini e animazioni, mappe immagine, componenti multimediali, collegamenti ipertestuali, grafici e schemi, componenti interattive, frame, tabelle e la verifica finale dell'accessibilità della pagina.

Per meglio comprendere finalità e obiettivi della Legge 9 gennaio 2004 n. 4, più comunemente nota come "Legge Stanca" o "Legge sull'accessibilità", occorre richiamare alcuni aspetti che caratterizzano la nostra attuale società. Una prima considerazione riguarda la natura stessa della Pubblica Amministrazione, Centrale e Locale, così come oggi percepita. Essa viene assimilata, nel comune sentire, ad un unico gigantesco apparato che, pur in una diversificata serie di articolazioni specializzate, ha come suo compito principale l'erogazione dei servizi necessari alla vita sociale e per questo scopo è sovvenzionata con quella sorta di colletta generale che si realizza attraverso le imposte. In altre parole, tutti i cittadini pagano per ottenere, sia a livello nazionale sia a livello territoriale, i principali servizi di cui hanno bisogno (sicurezza, formazione, assistenza, certezza del diritto).

Accade ora che la Pubblica Amministrazione, per razionalizzare la propria organizzazione, contenere i costi e migliorare al tempo stesso il livello di servizio (tempestività, capillarità,

affidabilità) debba ricorrere in misura sempre maggiore alle Tecnologie dell'Informazione e della Comunicazione (ICT), ed in particolare alle reti informatiche. Appurato che, per motivi tecnici ed economici, le reti internet ed intranet costituiscono, e sempre più costituiranno, lo strumento principe per offrire adeguati servizi ai cittadini occorre però porsi l'obiettivo di poter soddisfare tutta la potenziale utenza e non solo una sua quota parte. Se 57 milioni di persone pagano per ottenere servizi, la validità della soluzione proposta si misura con la percentuale di individui realmente serviti: una risposta è accettabile se tiene conto delle esigenze dei cittadini e del contributo di attività e di conoscenze loro richiesto.

Non avrebbe senso portare in tutte le case un terminale, potenzialmente in grado di fornire i più svariati servizi, senza preoccuparsi della semplicità d'uso delle applicazioni e della padronanza che l'abitante di quella casa può vantare sullo strumento. Anche con le più moderne tecnologie occorre soddisfare due condizioni: deve essere consentito, e se possibile agevolato, l'accesso alle informazioni per poter esprimere il proprio fabbisogno di servizio e, successivamente, devono essere chiare le modalità con cui muoversi tra le alternative proposte e le informazioni richieste ovvero si deve poter "navigare" senza problemi nell'applicazione. Vi è dunque un'esigenza di accessibilità che si estende immediatamente nel concetto di fruibilità del servizio. Ciò vale in particolare per quei cittadini che, in quanto disabili, possono con modesti accorgimenti tecnici avvalersi delle opportunità offerte dalle ICT dalle quali, per negligente disattenzione, potrebbero essere emarginati, senza peraltro poter disporre di valide alternative.

Il filo logico della legge Stanca può dunque riassumersi nei seguenti termini: la PA deve erogare servizi e questi saranno sempre più forniti attraverso reti informatiche; detti servizi devono essere resi disponibili al maggior numero di utenti e quindi anche a quel 5% di cittadini italiani portatori di qualche disabilità.

Occorre precisare che la legge Stanca non si limita a combattere forme di emarginazione a scapito del cittadino disabile regolamentando i livelli minimi di accessibilità dei siti web e dei servizi gestiti dalla PA. Poiché l'informatica può essere vista non solo come strumento di formazione e di informazione ma anche come strumento di lavoro, sono anche previste misure specifiche volte ad evitare discriminazioni nei confronti dei lavoratori disabili che, a norma di legge, almeno nelle Amministrazioni e nelle aziende più grandi, dovrebbero già rappresentare il 7% dei dipendenti. Al riguardo la legge recita testualmente: "I datori di lavoro pubblici e privati pongono a disposizione del dipendente disabile la strumentazione hardware

e software e la tecnologia assistiva adeguata alla specifica disabilità, anche in caso di telelavoro, in relazione alle mansioni effettivamente svolte”.

La legge Stanca, che sul tema si colloca all'avanguardia a livello internazionale per contenuti e visione d'insieme, affronta l'accessibilità partendo dalla Pubblica Amministrazione in ottica di gradualità di un intervento che è destinato ad interessare l'intera comunità. Si comincia pertanto dal più grande erogatore di informazioni e servizi, che è al tempo stesso il più grande datore di lavoro; il mondo privato, come già accaduto altre volte, non tarderà ad allinearsi agli standard ai quali i cittadini si saranno abituati.

È importante notare che, anche se la legge fa riferimento a “Disposizioni per favorire l'accesso dei soggetti disabili agli strumenti informatici”, in realtà viene promossa una cultura dell'attenzione verso tutte le categorie deboli, tra le quali la più numerosa e gravida di implicazioni socio-economiche è quella degli anziani.

In conclusione la legge va invece letta come un primo passo verso una più ampia strategia dell'inclusione (in opposizione quindi al Digital Divide) che, per gli aspetti etici, economici e sociali, rappresenta la miglior soluzione per una società organizzata che voglia definirsi civile.

1.3.1. Limiti della legge Stanca

Un disegno di legge approvato all'unisono da maggioranza e opposizione non è cosa di tutti i giorni, coi tempi che corrono. Ma è successo per la legge che si intitola “Disposizioni per favorire l'accesso dei soggetti disabili agli strumenti informatici”.

Gli obblighi introdotti dalla legge riguardano le pubbliche amministrazioni (art. 3 e 4), con precise disposizioni anche in materia di appalti. I privati (per i quali si prevedono contributi, ma non si stanziavano fondi ad hoc) sono obbligati a mettere a disposizione dei dipendenti disabili “la strumentazione hardware e software e la tecnologia assistiva adeguata alla specifica disabilità, anche in caso di telelavoro, in relazione alle mansioni effettivamente svolte”.

Lo stesso obbligo è imposto alle pubbliche amministrazioni ma, ahinoi, “nell'ambito delle disponibilità di bilancio”, e riguardano anche il “materiale formativo e didattico utilizzato nelle scuole di ogni ordine e grado”.

Un aspetto particolarmente importante, che segna forse una svolta nel processo di innovazione della pubblica amministrazione, è la previsione di responsabilità e sanzioni per i dirigenti e i dipendenti che non rispettino le disposizioni della legge (art. 9).

Comunque l'impatto effettivo del provvedimento potrà essere valutato solo dopo l'emanazione di un regolamento di attuazione da parte del Governo e di un regolamento tecnico da parte del Dipartimento per l'innovazione (art. 10 e 11). Tocca dunque soprattutto al ministro Stanca, che ha fortemente voluto questa legge, trasformare le dichiarazioni di principio in disposizioni efficaci, nei limiti consentiti dal testo varato dalle Camere.

Limiti che appaiono evidenti a una lettura attenta dell'articolato. Solo per citarne uno, si parla di hardware, software e tecnologie assistive, ma si dimentica che ci sono tanti disabili motori che non hanno bisogno di siti "accessibili", ma di poter "accedere ai siti", cioè di avere a condizioni favorevoli un collegamento a larga banda, che per molti di loro è l'unica possibilità di lavorare.

Non dovrebbe essere difficile escogitare una norma che preveda, per esempio, qualche agevolazione per i provider che vogliano destinare una modestissima percentuale degli abbonamenti alla fornitura gratuita di accessi ai disabili motori. Uno piccolo sforzo per risolvere un grande problema.

1.4. Limiti ed evoluzioni degli standard attuali

I limiti degli standard attuali sono essenzialmente dovuti alle carenze legate all'accessibilità al web da parte di individui con disabilità diverse da quelle visive. Nel documento proposto dalla WAI "Web Content Accessibility Guidelines 1.0" mancano riferimenti espliciti alle misure atte a favorire l'accesso a Internet per persone con disabilità per esempio cognitive.

Con il documento in fase di sviluppo "Web Content Accessibility Guidelines 2.0" (working draft nel dicembre 2003), la WAI evidenzia una certa sensibilità a questo aspetto: in più occasioni vengono suggeriti accorgimenti specifici per favorire la fruizione dei contenuti da parte di questa categoria di disabili, spesso trascurata sul profilo normativo.

In particolare si fa riferimento ad accorgimenti quali descrizioni grafiche alternative alla classica presentazione dei contenuti in forma testuale. Osservando la situazione attuale del web, ci si accorge che la possibilità di presentare i contenuti in forma di immagini o animazioni è stata finora solo considerata da pochissimi siti. Le norme in via di sviluppo suggeriscono anche di evitare termini e concetti inutilmente complicati e di favorire la navigabilità mettendo a disposizione strumenti come la mappa del sito, la tabella dei contenuti, il glossario, ecc.

La WCAG 2.0 prevede che le linee guida per la progettazione di contenuti web sia (citazione testuale):

1. Perceivable. Ensure that all content can be presented in form(s) that can be perceived by any user - except those aspects of the content that cannot be expressed in words.
2. Operable. Ensure that the interface elements in the content are operable by any user.
3. Understandable. Make it as easy as possible to understand the content and controls.
4. Robust. Use Web technologies that maximize the ability of the content to work with current and future accessibility technologies and user agents.

In particolare il checkpoint 1.5 [Extended] parla di navigazione semplificata e prevede che "Structure has been made perceivable to more people through presentation(s), positioning, and labels.". Inoltre il checkpoint 3.3 dice "Content is written to be no more complex than is necessary and/or supplement with simpler forms of the content". In particolare si parla di vocabolari controllati (Controlled languages use a restricted vocabulary taken from natural language. The purpose is to make texts easier to understand and translate. Standards generally limit words to a single meaning and prescribed part of speech. Complex syntax is avoided) e dell'uso di alternative non-testuali al testo e di supporto alla conversione in linguaggi simbolici ("providing support for conversion into symbolic languages" e "adding non-text content to the site for key pages or sections specifically to make the site more understandable by users who cannot understand the text only version of the site").

1.4.1. Il rapporto WWAAC

Il progetto WWAAC, acronimo di "World Wide Augmentative and Alternative Communication", è un progetto finanziato dalla comunità europea che ha l'obiettivo di rendere Internet e l'uso della posta elettronica più accessibili alle persone che usano la Comunicazione Aumentativa Alternativa. In questo progetto è stata condotta un'indagine per valutare lo stato dell'arte in fatto di accessibilità da parte di disabili che usano strumenti di CAA al web e ad altri moderni mezzi di comunicazione quali e-mail, forum di discussione e chat. Il progetto dovrebbe produrre nei prossimi mesi un browser studiato per gli utenti di CAA.

Uno dei primi risultati del progetto è una ricerca, chiamata "User Requirements Document", scritta da Mike Clark dell'ACE Center (<http://www.ace-centre.org.uk/>), che è il risultato di una serie di interviste con utenti di CAA, operatori che lavorano con strumenti di CAA e service providers. È stata fatta una panoramica dettagliata dei prodotti software attualmente

disponibili in commercio, cercando di mettere in luce le diverse barriere ancora presenti. I risultati sono molto interessanti e si possono raggruppare in 5 requisiti espressi dagli utenti di CAA e dagli operatori, elencati in ordine di priorità:

1. Accesso a contenuti e informazioni presenti sul web;
2. Possibilità di avere un servizio di posta elettronica tra utenti che parlano la stessa lingua (inglese-inglese; svedese-svedese...);
3. Utilizzo di forum di discussione;
4. Uso di posta elettronica tra utenti che parlano lingue diverse (inglese-svedese, ecc.);
5. Servizi di Chat.

Un altro risultato estremamente importante dello studio di Mike Clark è quello relativo alle motivazioni all'uso di Internet da parte degli utenti. Infatti tutti gli utenti hanno mostrato un grande interesse verso Internet, anche e soprattutto quelli di loro più penalizzati dalla situazione attuale. Questo vale sia per i servizi informativi che per i servizi relazionali e di community (forum, e-mail, chat).

Bisogna infine ricordare il fatto che, in generale, il target della Comunicazione Aumentativa Alternativa non è esclusivamente costituito da persone con disabilità cognitive: anche un bambino può usufruire più efficacemente dei contenuti se questi sono presentati con una forma a lui nota come ad esempio attraverso icone o figure.

1.5. Applicazioni degli standard

Nel seguito vengono presentati alcuni esempi di siti che applicano gli standard sull'accessibilità, sia italiani che stranieri. I siti sono raggruppati in "applicazioni standard di base", ossia quelli che applicano bene il livello base degli standard, e "applicazioni avanzate", ossia i siti che applicano concetti avanzati presenti nella normativa o ancora non codificati nella stessa.

Per ogni sito sono descritte brevemente le caratteristiche di accessibilità. Esula dagli scopi di questo paragrafo una trattazione approfondita degli obiettivi, degli utenti target e dei contenuti dei siti.

Applicazione standard di base

Presidenza Italiana del Consiglio dell'Unione Europea

(<http://www.ueitalia2003.it/ITA/>)

Il portale vanta le certificazioni HTML 4.01, CSS e WAI-A.

Agenzia delle entrate (<http://www.agenziaentrate.it>)

La caratteristica più importante di questo sito è la possibilità della personalizzazione delle modalità di presentazione. Accedendo al menù di personalizzazione (a dire il vero un po' nascosto) si può modificare la dimensione del carattere e soprattutto si può scegliere tra 3 tipi di presentazioni: standard, solo testo e con icone grandi.

Portale del governo italiano (<http://www.italia.gov.it>)

C'è la possibilità di sfogliare il sito attraverso un portale vocale, disponibile chiamando un numero verde. Sono presenti le certificazioni più rilevanti: HTML 4.01, CSS e WAI-A.

Portale sulla disabilità australiano (<http://www.accessibility.com.au>)

Sito molto interessante poiché ben strutturato e con tutte le certificazioni rilevanti (Bobby Approved-AAA, Section 508, HTML 4.01, CSS).

Camera dei deputati (<http://www.camera.it>)

Le particolarità del sito sono la presenza di una versione solo testo e la presenza di una versione altamente accessibile (<http://wai.camera.it>), la quale presenta un elevato contrasto cromatico, e caratteri abbastanza grandi, ma non sono presenti utilità per la personalizzazione.

Applicazioni avanzate

SuperAbile (<http://www.superabile.it>)

C'è la possibilità di consultare il portale vocale telefonando ad un opportuno numero verde. Il portale permette la registrazione degli utenti e durante la profilazione viene chiesto a quale tipo di versione si vuole accedere; sono disponibili tre versioni di personalizzazione della presentazione: versione standard, solo testo oppure versione ad alta leggibilità. Viene richiesto anche l'eventuale utilizzo di ausili nell'utilizzo del PC, oltre che altri dettagli della

personalizzazione quali risoluzione dello schermo e velocità di connessione. Il sito è accessibile con livello WAI-A.



Figura 1: superabile.it in versione ad alta leggibilità

Portale per disabili (<http://www.visionconnection.org>)

Questo portale ha un servizio di registrazione e profilazione che permette una presentazione personalizzata dei contenuti. Si può impostare la dimensione del carattere e il contrasto cromatico.

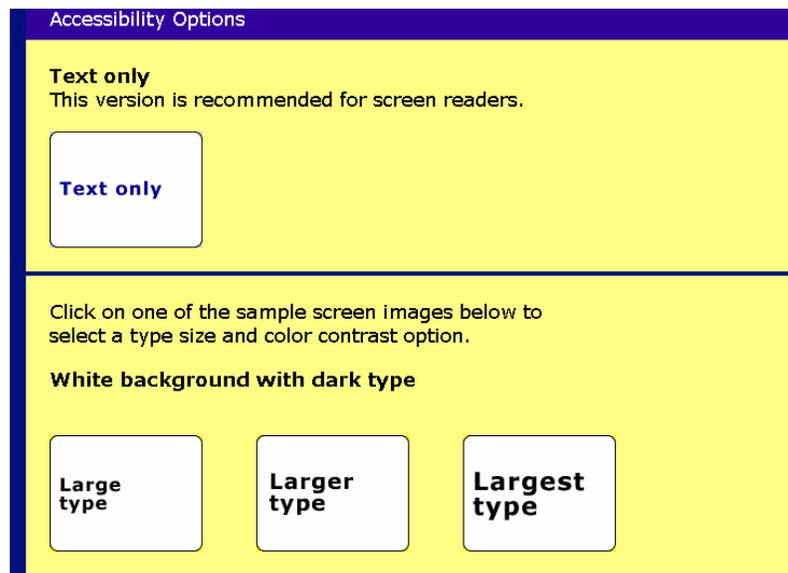


Figura 2: Pagina di personalizzazione del portale VisionConnection

Eraclito (<http://accessibilita.eraclito.it>)

La Eraclito sviluppa da anni, applicazioni, prodotti e servizi web accessibili in stretta collaborazione con la maggiore delle associazioni di non vedenti: L'Unione Italiana dei Ciechi. Nel sito di Eraclito è presente una pagina di personalizzazione della presentazione dei contenuti: è possibile scegliere il tema e impostare il carattere.

Unione Italiana Ciechi (<http://www.uiciechi.it>)

La particolarità di questo sito è il servizio di personalizzazione, che permette l'impostazione del contrasto e la piena configurabilità dei font; sono inoltre presenti quattro configurazioni di default, a cui corrispondono quattro modalità di presentazione dei contenuti. Il sito è certificato WAI con il massimo livello di accessibilità. Sul sito è presente una sezione dedicata all'Osservatorio Siti Internet (OSI), che è un gruppo di circa una trentina di disabili, sensibili a queste tematiche. Tale gruppo ha avuto collaborazioni importanti per rendere accessibili alcuni siti appartenenti alla pubblica amministrazione.

Puoi scegliere una delle configurazioni standard oppure personalizzare ogni singolo elemento. Potrai vedere il risultato delle modifiche nell'anteprima presente in fondo a questa pagina.

Configurazioni standard

Normale

[Torna alla pagina iniziale](#)

[Personalizza il tipo di carattere](#)

[Personalizza il colore dello sfondo e delle indicazioni](#)

[Personalizza l'allineamento del testo](#)

Figura 3: Menù di personalizzazione del sito dell'Unione Italiana Ciechi

Peepo (www.peepo.com)

Portale per disabili cognitivi con navigazione attraverso icone PCS.

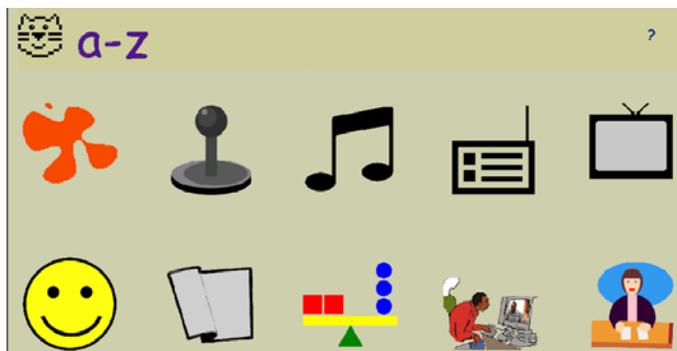


Figura 4: Homepage di peepo.com

Papunet (www.papunet.net)

Portale finlandese per disabili cognitivi con alcune sezioni realizzate con simboli Bliss.



Figura 5: Una pagina di contenuto del portale papunet.net

1.6. Impatto dell'accessibilità sui costi e sulle modalità di progettazione delle architetture web: project management per l'accessibilità

Nell'ultimo decennio, con l'introduzione di sempre nuove tecnologie comunicative (animazioni flash, giochi interattivi con applet java...), i webmaster hanno spinto per soluzioni grafiche sempre più "estreme", privilegiando presentazioni ad affetto piuttosto che usabilità e accessibilità.

In questi ultimi anni, tuttavia, si è affiancata una nuova tendenza, quella del "design for all", cioè quella della progettazione universale, che tenga quindi conto delle molteplici esigenze corrispondenti alla disomogeneità dell'utenza. Storicamente i primi siti ad accorgersi di questa

mancanza sono stati quelli preposti alle vendite online; si accorsero del numero enorme di potenziali clienti che restavano esclusi dall'offerta.

Nell'ambito della gestione di un progetto di accessibilità di grandi dimensioni, sono molteplici gli aspetti che vanno tenuti in considerazione: allocazione opportuna delle risorse, monitoraggio dello stato di avanzamento del sito, formazione dei dipendenti, individuazione delle aree di potenziali problemi e un'opportuna integrazione dell'accessibilità con la presentazione dei contenuti.

La maggior parte delle organizzazioni governative hanno un loro team di sviluppo web e internamente si possono individuare unità con responsabilità differenti (es. studio grafico, application developers, database developers, sistemisti); è importante ricordare che tutti queste unità sono coinvolte nel progetto, e devono condividere le tematiche inerenti all'accessibilità.

A titolo esemplificativo possiamo considerare un team composto da 5 unità (tutte partecipano attivamente al progetto):

1. Sviluppo contenuti: sono i redattori del sito / portale.
2. Sviluppo elementi multimediali: si tratta di figure preposte alla creazione di supporti multimediali richiesti da alcuni tipi di disabilità, come ad esempio l'help vocale.
3. Sviluppo del layout e design: si tratta tipicamente di esperti di usabilità che progettano la navigazione del sito e la sua strutturazione (dimensione orizzontale, profondità...).
4. Gestione del portale: sono gli sviluppatori dell'architettura software.
5. Studio grafico: è responsabile della veste grafica del sito.

Fondamentalmente possiamo individuare quattro macro fasi (Yonaitis, [24]) nell'ambito di un progetto di un sito accessibile (sia nel caso si tratti di un progetto di un nuovo portale, sia si tratti di un progetto di manutenzione di un sito esistente, volto ad ottenere certi obiettivi di accessibilità):

1. Sviluppo di un piano: definizione della strategia per rendere il sito accessibile; definizione del livello di accessibilità che si vuole raggiungere (ad esempio, con riferimento alle norme WAI, possiamo raggiungere il livello A, AA oppure AAA, corrispondente alla massima accessibilità); definizione degli obiettivi dell'organizzazione; valutazione dello stato attuale del sito.

-
2. Applicazione delle linee guida per l'accessibilità; viene di solito eseguita con appositi tools di validazione e riparazione: si tratta di programmi che esaminano pagine web, e ne verificano la conformità con un certo set di norme. Qualora la pagina non fosse accessibile, vengono suggerite delle modifiche per renderla accessibile. In questa fase vengono trasmessi dei feedback alla fase di sviluppo della strategia.
 3. Testing della versione sperimentale del sito.
 4. Manutenzione: si tratta di una fase spesso tralasciata, e la sua trascuratezza ha portato al fallimento di numerosi progetti di accessibilità. Gli aspetti inerenti alla manutenzione devono essere considerati anche e soprattutto in fase progettuale. Questa fase fa comprendere come sia indispensabile fare attenzione alla conformità alle norme sull'accessibilità, anche dopo che il portale è stato rilasciato.

E' importante tenere presenti durante tutte le fasi gli obiettivi di accessibilità che sono stati definiti.

Per un piccolo sito è possibile analizzare singolarmente ogni pagina ed apportare le modifiche che sono necessarie per raggiungere il livello di accessibilità che si è fissato.

Per un portale di grosse dimensioni questo approccio non è praticabile, anche e soprattutto perché la manutenzione sarebbe troppo onerosa.

In queste situazioni è necessario reingegnerizzare il portale. In questa fase i tools di validazione automatica possono fornire in poco tempo un quadro della situazione attuale del sito.

Tenuto conto della situazione attuale, si può definire il livello di accessibilità richiesto. Altri evidenti vincoli da tenere in considerazione in questa scelta sono le risorse disponibili e l'esperienza maturata precedentemente in questo campo.

A questo punto si possono schedare le attività per raggiungere gli obiettivi fissati.

Nel caso di portali di grosse dimensioni (come quelle della Pubblica Amministrazione) con contenuti che vengono aggiornati frequentemente, devono essere sviluppati dei componenti redazionali a supporto del meccanismo di pubblicazioni.

In particolare, l'ambiente di back office deve essere arricchito con servizi aggiuntivi per la validazione dei nuovi contenuti. Il Content Management System (CMS: è un componente software di back-end che serve per supportare l'aggiornamento dei contenuti) può essere integrato con le suddette funzioni. Ad esempio, quando il redattore inserisce un'immagine,

viene automaticamente richiesto di inserire il testo alternativo all'immagine, permettendo ai non vedenti di accedere comunque all'informazione, anche se parziale.

Nell'ambito dello sviluppo di elementi multimediali per particolari categorie di disabili è opportuno prevedere un'adeguata formazione del personale con particolare riferimento alle tematiche riguardanti l'accessibilità; sono infatti notevoli le norme WAI che regolamentano l'utilizzo di questi elementi.

Diversi casi aziendali dimostrano che, vista la complessità di progetti di questo tipo (molto legata all'elevato numero di attori che sono interessati), è consigliabile stabilire inizialmente degli obiettivi parziali rispetto alle esigenze reali. Questo si traduce concretamente nel porsi come obiettivo, ad esempio, un livello di accessibilità di tipo "A" piuttosto che "AAA", oppure è preferibile, almeno inizialmente, rendere accessibili solo alcune sezioni del portale, e solo in una seconda fase estendere le modifiche (che sono spesso strutturali) al resto dell'architettura.

Tecniche formali come la **pianificazione e il controllo dello stato di avanzamento dei lavori** sono fondamentali in quanto si ha a che fare con progetti di grosse dimensioni.

La presenza di diversi attori, con competenze molto differenti, rende fondamentale altresì l'impiego di metodi di **integrazione esterna**, atti a favorire l'integrazione del team di sviluppo dell'architettura con gli esperti del dominio. Questi metodi prevedono accorgimenti quali: frequenti riunioni di diversi rappresentanti dei vari attori, approvazione formale delle specifiche e delle eventuali loro modifiche, scelta di un esperto del dominio come responsabile del progetto.

In generale si può dire che le **grosse dimensioni** del progetto, unitamente alla **non strutturazione** dello stesso, fanno sì che i progetti legati all'accessibilità sono spesso **rischiosi**, e i metodi proposti sono proprio volti a diminuire le eventuali conseguenze del rischio: parziale o mancato raggiungimento degli obiettivi, costi e/o tempi superiori a quelli preventivati.

2. Introduzione al progetto “navigAbile©”

“navigAbile” è un progetto della Fondazione Italiana Accenture - in partnership con la Fondazione Don Carlo Gnocchi e la Fondazione Francesca che intende fornire, principalmente a bambini e ragazzi con disabilità motorie e di comunicazione, una modalità innovativa per accedere ai contenuti del web.

Ho partecipato a questo progetto nell’ambito di uno stage che ho svolto presso la società di consulenza Accenture, che è iniziato nel Settembre 2003 e si è concluso nel Giugno 2004; in questa attività sono stato impegnato part-time, durante l’ultimo anno di università.

Il progetto “navigAbile” ha due principi guida: **l’utente di internet è abile in n modi diversi**, al di là di considerazioni sulla sua eventuale disabilità o meno, e **l’utente di internet ha bisogno di essere abilitato in n modi diversi**, a seconda di quale tipo particolare di abilità gli è più congeniale. L’abilitazione avviene sia attraverso gli ausili utilizzati dagli utenti (e.g. tastiera e mouse speciali), sia attraverso le caratteristiche di accessibilità intrinseche dei siti o dei portali internet. È indubbio che entrambi gli aspetti siano essenziali. **Tuttavia il focus del progetto navigAbile è sul miglioramento delle caratteristiche di accessibilità dei siti e non sugli ausili tecnologici utilizzati dagli utenti per interagire con il computer.**



Figura 6: Il logo di navigAbile©: tutti i diritti riservati

Nel contesto attuale, il tema dell’accessibilità del web per tutti sta diventando sempre più centrale. Lo stesso Tim Berners-Lee, padre del web, ha ribadito recentemente che “Il potere del web sta nella sua universalità. Un aspetto essenziale è che tutti possano accedervi a prescindere dalle disabilità”. Le iniziative volte ad eliminare le “barriere elettroniche” si moltiplicano, sia in Europa (il 2003 è stato l’Anno Europeo delle Persone con Disabilità) che in altri paesi, quali gli Stati Uniti, l’Australia, il Canada. In particolare l’aderenza agli standard del Rehabilitation Act Americano (la famosa sezione 508) e del W3C, il consorzio internazionale presieduto da Tim Berners-Lee, che è anche il promotore della Web Accessibility Initiative (WAI), è diventato un requisito irrinunciabile per gli enti pubblici e per moltissime aziende. Lo scorso Giugno il ministro Stanca ha dichiarato che “Entro tre anni dall’approvazione del

disegno di legge del governo per favorire l'accesso dei soggetti disabili agli strumenti informatici tutti i siti delle pubbliche amministrazioni centrali saranno accessibili". La spinta c'è ed è evidente. Le motivazioni possono essere le più diverse, da ragioni etiche e sociali a quelle più "commerciali" di aziende che si rendono conto che, trascurando il tema dell'accessibilità, stanno letteralmente chiudendo la porta in faccia a milioni di utenti potenziali. La stima della percentuale di disabili sulla popolazione va dal 5% al 15%, a seconda dei parametri di valutazione usati nelle statistiche. In ogni caso anche il 5% della popolazione si concretizza in 2,8 milioni di Italiani!

La vision del progetto navigAbile è quella di contribuire a realizzare un web diverso:

- ◆ Un web dove l'accessibilità sostanziale e non formale sia considerata una priorità irrinunciabile per ogni sito o portale.
- ◆ Un web che possa essere uno strumento di formazione e di crescita per tutti, in particolare per i disabili in età evolutiva, perché possa diventare poi uno strumento di lavoro e di vita quotidiana, e non un ulteriore ostacolo da superare.
- ◆ Un web a cui la persona disabile possa accedere in modo semplice e intuitivo, senza la necessità di dover configurare e padroneggiare complessi ausili software.
- ◆ Un web che si adatti alle esigenze di accessibilità/navigabilità dell'utente e non viceversa, all'insegna dell' "accessibilità per tutti, ma su misura per ciascuno".

La realizzazione di questa vision richiede di agire su tre livelli distinti: il livello "culturale", il livello degli ausili e degli strumenti hardware e software necessari per collegarsi a internet e il livello dei siti/portali presenti sul web. Tuttavia oggi vi è una "cultura web" diffusa (soprattutto tra i giovani anche disabili) e il settore degli ausili ha raggiunto una ragionevole maturità. Sono invece molto spesso bloccanti le barriere che riguardano il cuore del "sistema web", ossia le caratteristiche di accessibilità di siti e portali, come mostrato nell'immagine seguente:

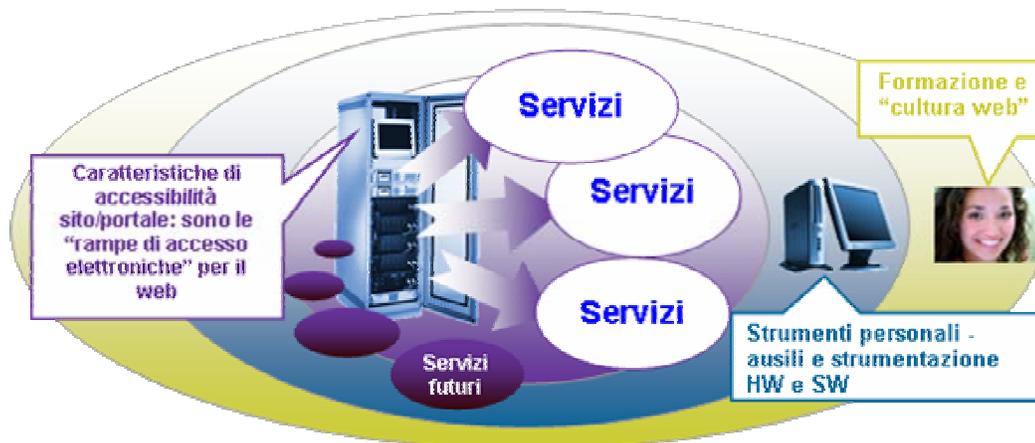


Figura 7: Schema che evidenzia il Focus del progetto navigAbile

Per migliorare il livello di accessibilità del web bisogna mettere in campo delle iniziative che vadano al di là dei pur opportuni interventi normativi e di standardizzazione. È necessario realizzare componenti architettonici e servizi riutilizzabili da più soggetti, diminuendo in questo modo il "costo dell'accessibilità" per chi sviluppa siti e portali. Inoltre, come sarà spiegato nel paragrafo sul "Contesto" parlando dei limiti degli standard e delle loro applicazioni, sappiamo che, anche se gli standard più autorevoli venissero applicati completamente, ci sono intere categorie di disabili che troverebbero comunque barriere spesso insormontabili per accedere al web. Questo è tanto più drammatico se si pensa che un ragazzo disabile, escluso dal web in età evolutiva, non avrà modo di sviluppare appieno le proprie potenzialità ed abilità nell'utilizzare quello che è e sarà sempre più uno strumento indispensabile nella nostra vita privata e lavorativa.

Da queste considerazioni discende la **mission** del progetto navigAbile:

"Progettare, sviluppare e validare scientificamente un'architettura di servizi abilitanti per migliorare l'accessibilità del web prioritariamente alle persone con disabilità in età evolutiva che utilizzano strumenti di comunicazione multimodale (e.g. Comunicazione Aumentativa Alternativa, ecc.), permettendo:

- 🌈 di recuperare, valorizzare e potenziare le diverse abilità possedute dagli utenti, con particolare riferimento alle abilità relazionali e di comunicazione;*
- 🌈 di rendere disponibile un'interfaccia multimodale in grado di adattarsi alle esigenze e al profilo del singolo utente (e non viceversa);*
- 🌈 di permettere agli utenti con disabilità di comunicazione e relazione di poter accedere, sia in modo diretto che attraverso la mediazione di un'altra persona, agli stessi*

contenuti e servizi proposti ai loro coetanei (contenuti formativi, contenuti informativi, servizi di comunicazione e relazionali ecc.), favorendo l'integrazione degli utenti sul web;

- 🚦 di facilitare l'abbattimento per le persone con disabilità delle "barriere elettroniche" del web (necessità di supporto e costi per download software ad hoc, rigidità nell'uso "solo" del proprio pc, ecc.);*
- 🚦 di mettere a disposizione di chi produce siti e contenuti sul web una serie di servizi architettonici che permettano di abbassare le "barriere di costo" per lo sviluppo di un sito altamente accessibile;*

Il progetto, dopo questa fase di progettazione, sviluppo e validazione, potrà estendersi ampliando da un lato le modalità di comunicazione multimodali supportate e, dall'altro, fornendo i servizi architettonici sviluppati ad un numero sempre maggiore di siti, portali e relativi contenuti."

L'architettura di servizi navigabile è costituita da tre aree: **Servizi di registrazione e profilazione, Servizi di personalizzazione e navigazione, Servizi di supporto alla gestione dei contenuti**. L'ottica con cui sono pensati i servizi architettonici è quella di utilizzare le "best-practice" presenti nel mondo web e di applicarle al contesto dell'accessibilità, rispondendo anche alle esigenze di utenti non ancora considerati dagli stessi standard sull'accessibilità citati. **Non si tratta di reinventare nulla, ma di organizzare in un'architettura coerente, che ha come obiettivo la massima accessibilità del web, strumenti e servizi realizzati in modo disomogeneo in realtà diverse o magari applicati ad altri contesti**. L'architettura di servizi sviluppati potrà poi essere messa a disposizione di siti/portali per aumentarne l'accessibilità diminuendo i costi associati.

La realizzazione della soluzione proposta, considerata anche la complessità del contesto affrontato, richiede il coinvolgimento di **diverse entità** con ruoli complementari. Si è ritenuto tuttavia cruciale che, almeno in una prima fase, il numero di attori sia ridotto al minimo necessario per raggiungere gli obiettivi previsti, per evitare la dispersione di tempo ed energie nel lavoro di coordinamento di realtà diverse ed eterogenee. Oltre alla **Fondazione Italiana Accenture** (promotrice e finanziatrice principale del progetto) e ad **Accenture**, saranno quindi coinvolti la **Fondazione Don Gnocchi**, attiva da più di 50 anni nell'assistenza e nel recupero dei disabili, la **Fondazione Rava** e **una scuola pilota**. Oltre a questo nucleo

minimo di protagonisti, potranno essere aggregati altri partner tecnologici, istituzionali o scientifici con il ruolo di validazione e supporto al lavoro del gruppo principale di progetto.

La **pianificazione e la gestione delle attività di progetto** avverranno con una strategia di tipo iterativo, secondo le best-practice ormai riconosciute nel campo del project management, al fine di minimizzare i rischi di progetto e massimizzare l'efficacia dell'intervento. Ciò si concretizza con la divisione del progetto in fasi e la suddivisione della fasi in iterazioni, secondo la logica del **"think big, start small, scale fast"** (pensa in grande, parti in piccolo, cresci velocemente). Nella prima fase verrà progettata l'architettura generale (think big) e realizzate le sue componenti di base (start small), in modo da essere pronti ad espandere l'orizzonte del progetto nella fase 2, dopo la verifica degli esiti della fase 1, con il coinvolgimento di nuovi attori e con la realizzazione di nuovi servizi (scale fast).

2.1. La Soluzione proposta

Il progetto navigAbile, nella prima fase conclusasi nel marzo 2004, ha realizzato diversi prodotti:

1. **Servizi architetturali:** navigAbile è innanzitutto un'architettura orientata ai servizi (detta anche SOA o *Service Oriented Architecture*). I servizi architetturali sono quindi il primo prodotto del progetto e sono le "fondamenta" di navigAbile, sia per le applicazioni attuali (si veda sito dedicato) che per quelle future. Tra i servizi di base ricordiamo:
 - a. **Servizi di Interoperabilità:** sono l'interfaccia di navigAbile verso il mondo esterno. I servizi sono esposti come *web services*, in modo da poter essere richiamati da altri siti/portali che volessero rendersi 'NavigAbili';
 - b. **Profilazione:** servizio di registrazione dei profili utente e dei requisiti di personalizzazione associati (si veda la descrizione nel sito dedicato);
 - c. **Filtro sulle pagine (*Page Filtering*):** il sistema di *Page Filtering* integra in sé un meccanismo di costruzione delle pagine che permette di aumentarne l'accessibilità e di personalizzare i contenuti e la presentazione degli stessi. Il filtro utilizza sia un *tag system* proprietario navigAbile, che la struttura HTML della pagina;

-
- d. **Servizi di comunicazione:** sono i servizi utilizzati per facilitare la comunicazione, come per esempio il dizionario multidimensionale, grazie al quale ogni contenuto può essere associato a diversi formati (Testo, PCS, Bliss, Animazioni...), le tabelle di comunicazione, il compositore frasi. Per una descrizione più dettagliata si rimanda al paragrafo di descrizione del sito dedicato;
 - e. **Servizi di supporto alla creazione di contenuti accessibili:** questi servizi supportano la produzione di contenuti accessibili, tramite l'utilizzo di indicatori di complessità del testo e la verifica dell'appartenenza dei termini usati a un dizionario controllato.
2. **Sito dedicato:** il sito dedicato di navigAbile (raggiungibile all'indirizzo www.navigabile.it) è nato per tre esigenze diverse: sottoporre a test i servizi architetture, realizzando un primo esempio di applicazione che li mostrasse in azione; applicare i servizi navigAbile a un contesto semplice ma realistico di un sito di informazione; mettere a disposizione da subito agli utenti coinvolti nella sperimentazione alcuni servizi web accessibili, quali la bacheca o le news.

La realizzazione di questi prodotti ha portato anche alla definizione di un **metodo**, con cui sono stati identificati i cluster di utenti, definiti i requisiti, progettati, sottoposti a test, sperimentati e validati scientificamente i servizi di navigAbile.

Nel seguito verrà descritto il sito dedicato navigAbile. Attraverso la descrizione del sito verranno mostrati in azione quelli che sono i servizi architetture di navigAbile.

2.2. L'architettura dei Servizi di navigAbile: introduzione

L'architettura di servizi navigAbile si rivolge a tre aree di funzionalità normalmente presenti su un portale web:

- ◆ **Servizi di supporto alla gestione dei contenuti;**
- ◆ **Servizi di personalizzazione e navigazione;**
- ◆ **Servizi di registrazione e profilazione;**

L'ottica con cui sono descritti i servizi architeturali è quella di utilizzare le migliori soluzioni presenti nel mondo web e di applicarle al contesto dell'accessibilità. Molto di ciò che verrà descritto è presente, sviluppato in modo più o meno compiuto, sul web. Non si tratta di reinventare ciò che già c'è, ma di organizzare in un'architettura coerente, ed eventualmente di completare, strumenti e servizi presenti in modo disomogeneo in realtà diverse o magari applicati ad altri contesti. I servizi verranno presentati attraverso una descrizione del servizio stesso, così come viene normalmente incluso nelle architetture tradizionali, attraverso l'enunciazione del valore aggiunto per l'utente e le possibili evoluzioni che l'architettura navigabile può introdurre per aumentare l'accessibilità del web.

2.2.1. Servizi di supporto alla gestione dei contenuti

Descrizione generale: Il servizio di gestione dei contenuti permette di inserire, modificare e aggiornare i contenuti di un sito o portale. In genere il servizio permette di gestire anche utenti e ruoli (fornitori di contenuti, redattori e caporedattori) e il processo di "lavorazione" dei contenuti inseriti (inserimento in bozza, approvazione, pubblicazione).

Valore aggiunto per l'utente: Il vantaggio principale per l'utente è che la gestione strutturata di una redazione di tipo giornalistico per la produzione dei contenuti è la miglior garanzia che i contenuti presenti sul sito siano continuamente e dinamicamente aggiornati. Evidenti sono poi i vantaggi per gli utenti "interni", ossia coloro che hanno il compito di produrre i contenuti: un sistema di content management garantisce ai fornitori di contenuti, non necessariamente tecnici esperti di internet, la possibilità di avere a disposizione un ambiente che li supporti nelle loro attività.

Affinché un sito o un portale siano veramente accessibili, è fondamentale che i contenuti siano nella versione adatta all'utente. Una notizia o un capitolo di un corso di formazione devono essere descritti in modo diverso a seconda che l'interlocutore sia un adulto oppure un bimbo o ancora un adulto con difficoltà cognitive. Per taluni utenti un'animazione multimediale è una barriera insormontabile, per altri è un aiuto prezioso per comprendere i concetti presentati.

E' necessario quindi fornire delle versioni alternative semanticamente equivalenti dei contenuti. Alcuni esempi di alternative equivalenti:

- ◆ **Versione testuale standard:** è il contenuto presentato nella sua stesura originale;

-
- ◆ Versione testuale semplificata: il contenuto potrebbe essere presentato utilizzando un vocabolario limitato (VdB di T. De Mauro);
 - ◆ Versione vocale (testo letto da un lettore o generato da un sintetizzatore);
 - ◆ Versione multimediale/Animazione;

Oltre alla gestione di contenuti con versioni alternative, il sistema di CMS può supportare i fornitori di contenuti con:

- ◆ Vocabolario di base con alternative equivalenti: è il vocabolario di base limitato con cui sono composti i contenuti in versione semplificata. Per ogni termine del dizionario sono presenti delle alternative equivalenti (e.g. versione sonora, Bliss, PCS, animazione). Lo stesso dizionario, eventualmente con delle estensioni, può essere utilizzato anche nella presentazione dei contenuti come "glossario on-line";
- ◆ Supporto alla gestione di parole chiave (Keyword): per ogni contenuto, deve essere prevista la gestione delle keyword associate. Le Keyword devono appartenere al vocabolario di base, in modo che in fase di presentazione dei contenuti le stesse possano essere presentate in diverse alternative equivalente (Testo, PCS, Bliss, animazioni...)
- ◆ Indicatori di leggibilità del testo: usando formule di leggibilità presenti in letteratura (e.g. Gulpease), il redattore potrebbe venire forzato a rispettare dei livelli di leggibilità standard per alcune versioni equivalenti (e.g. rispetto all'indice Gulpease le soglie potrebbero essere 60% per la versione normale e 90% per la versione semplificata)

2.2.2. Servizi di Personalizzazione e Navigazione

Descrizione generale: I servizi di personalizzazione e di supporto alla navigazione includono sia la presentazione dei contenuti in modo coerente con quanto presente nel profilo di registrazione dell'utente, sia tutti quei "supporti" forniti all'utente per orientarsi e per navigare nei contenuti stessi.

Valore aggiunto per l'utente: Il valore aggiunto è una maggior usabilità o, come spesso si dice, una "user experience" di qualità superiore.

Possibili requisiti di massima dell'evoluzione navigabile: i servizi di personalizzazione e di navigazione contengono alcune componenti "classiche" mutate dalle architetture dei portali più diffuse, quali il motore di personalizzazione e vari supporti alla navigazione quali

motore di ricerca, servizio di directory, mappa del sito. Oltre a ciò dovranno però essere sviluppati componenti specifici per supportare l'utente nella navigazione e nella comprensione dei contenuti:

- ◆ Supporto alla fruizione dei contenuti (sui contenuti in versione testo o simbolica): rientrano in questa categoria tutti i supporti alla lettura del testo, ad esempio in una modalità che permetta ad esempio di evidenziare le parole una per volta, insieme alla possibilità di attivare la voce di un narratore, sintetica o no, che legga il testo (se la versione equivalente audio è disponibile);
- ◆ Supporto alla navigazione: lettura degli elementi di navigazione ogni volta che uno di questi riceve il focus e/o breve descrizione delle funzionalità (Help audio, tramite registrazioni o meccanismi di conversione "text-to-speech"). Per ogni elemento di navigazione è possibile, analogamente a quanto avviene per i contenuti, presentare l'equivalente non testuale in linguaggi alternativi (e.g. Bliss). Ciò richiede la gestione strutturata degli strumenti di navigazione;
- ◆ Visualizzazione delle parole chiave: come si è detto trattando del sistema di gestione dei contenuti, ad ogni contenuto saranno associate una o più parole chiave per guidare la comprensione. Queste parole chiave possono essere mostrate all'utente in linguaggio standard e in diversi linguaggi alternativi, quali ad esempio il Bliss o il PCS. La funzione delle parole chiave non è quella di sostituire il testo, ma di fornire una sorta di "orientamento" al testo prima di affrontarlo. Questa tecnica è sfruttata in diverse metodologie didattiche.

2.2.3. Servizio di Registrazione e profilazione

Descrizione generale: Il servizio di Registrazione e profilazione permette di acquisire e di gestire i dati relativi al profilo del singolo utente, che poi verranno utilizzati dal servizio di personalizzazione per "adattare" il sito o il portale alle reali esigenze dell'utilizzatore dei servizi.

Valore aggiunto per l'utente: in cambio del tempo "speso" per fornire informazioni relative al proprio profilo, l'utente deve percepire, durante l'utilizzo del portale, che queste informazioni sono sfruttate al meglio per garantirgli un livello di usabilità e di accessibilità più elevato.

2.2.4. La prima applicazione e la validazione scientifica

La prima applicazione dell'architettura navigAbile sarà effettuata su una sottosezione di un sito esistente. L'obiettivo della reingegnerizzazione del sito è quello di supportare lo Usability & Accessibility Test e di essere una "proof of concept" delle potenzialità dell'architettura, fino alla sua validazione scientifica insieme ai partner di progetto.

La scelta di reingegnerizzare una sezione di un sito esistente è funzionale agli obiettivi del progetto, che vanno verso la realizzazione di un'architettura da poter poi applicare in diversi contesti e su diversi siti.

Il sito o la parte di sito da reingegnerizzare deve presentare delle caratteristiche di accessibilità minime (Priorità 1 dei checkpoint WAI), deve essere dotato di un'infrastruttura per la registrazione/profilazione/personalizzazione e di un sistema di gestione dei contenuti (CMS).

2.3. Architettura dei servizi di navigAbile

Le soluzioni architettoniche a supporto del sistema navigAbile hanno l'obiettivo di realizzare un efficiente sistema di servizi rivolto a qualsiasi sito il cui proposito sia quello di migliorare l'accessibilità del web, con una particolare attenzione ai disabili in età evolutiva che utilizzano strumenti di comunicazione multimodale (per esempio la Comunicazione Aumentativa Alternativa, ecc.).

L'architettura si pone dunque i seguenti obiettivi di dettaglio:

- ◆ **Non ghettizzazione.** Il sito ospite, che integra opportunamente i servizi navigAbile, è nelle condizioni di offrire **più versioni alternative** non solo dei contenuti ma anche dei layout in cui questi sono presentati. Solo per le pagine di navigazione si ricorre a un template, comunque in linea con lo stile delle pagine e per questo a carico del sito stesso: per tutto il resto l'architettura integrata di navigAbile consente una personalizzazione legata alle esigenze e al gusto del singolo utente. Si evita così la ghettizzazione di particolari tipologie d'utenza a delle pagine pensate ad hoc, predisponendo invece il sito ad adattarsi al meglio a chi deve navigare le pagine e fruire dei contenuti/servizi offerti.
- ◆ **Servizi condivisibili.** Nella logica dell'accessibilità e della semplicità d'integrazione nelle condizioni più disparate, tutti i servizi di navigAbile sono concepiti per essere condivisibili a prescindere dalle tecnologie, dai linguaggi e dalle piattaforme: si tratta

infatti di servizi, estensibili a web services veri e propri, che fanno uso di protocolli "aperti" (SOAP per realizzare RPC) per essere svincolati in ogni modo da qualsiasi tecnologia in particolare.

- ◆ **Non invasività.** Le soluzioni presentate per l'integrazione dei servizi di navigAbile hanno la peculiarità di non essere invasive per il sito ospite: concettualmente, infatti, si collocano come uno strato software tra utente e sito. Questo meccanismo può essere realizzato fisicamente in diversi punti dello scenario in cui gli attori sono gli utenti, i siti internet e i servizi di navigAbile. Obiettivo primo dell'architettura è quello di creare un'infrastruttura integrata ma autonoma, caratterizzata da semplicità d'integrazione, che consenta di rendere il sito veramente accessibile, senza dover creare nuove pagine o navigazioni parallele, e senza perdere l'impronta e lo stile originale.
- ◆ **Semplicità d'uso.** Particolare attenzione è stata prestata, nella prima fase del progetto, all'obiettivo imprescindibile di creare delle soluzioni caratterizzate da una grande facilità d'utilizzo: a tale fine è stato ideato un sistema di *tag* (o marcatori) capace di arricchire pagine web preesistenti in modo da preparare tali pagine per un successivo filtraggio che costruisce e adatta le pagine per l'utente finale in modo dinamico. La Redazione del sito (a cui non è richiesta la conoscenza del linguaggio HTML) si limita quindi ad aggiungere alle pagine originali dei tag attraverso cui è possibile semplificare la pagina o indicare contenuti alternativi. Tale soluzione si appoggia dunque su ciò che già esiste, senza risultare invasiva.
- ◆ **Profilazione e personalizzazione.** I servizi e l'architettura navigAbile garantiscono la personalizzazione dei contenuti e della struttura delle pagine per garantire un'esperienza di navigazione vicina alle peculiari esigenze di ciascun utente. Il meccanismo di personalizzazione è legato a un profilo, registrato dall'utente stesso, che racchiude tutte le informazioni necessarie per consentire al sistema di adattare il più possibile i contenuti e la presentazione alle caratteristiche dell'individuo.
- ◆ **Flessibilità.** La flessibilità è tra gli importanti obiettivi dell'architettura esposta nel prossimo capitolo. Tale architettura deve essere infatti modulare, distribuita, scalabile e astratta tanto da consentire successive implementazioni di requisiti futuri. I livelli logici sono ben compartimentati, nel rispetto delle linee guida delle moderne architetture distribuite.

Il **primo proposito dell'architettura** è fornire servizi e soluzioni per rendere un qualsiasi sito in grado di rivolgersi a una platea più ampia ed eterogenea. Inoltre le soluzioni proposte non vogliono essere invasive per i siti potenzialmente interessati. I servizi sono, infatti, esposti come interfacce basate su SOAP, non indirizzate a una particolare tecnologia.

Se il generico sito è un edificio, questa architettura non si propone di rivoluzionare la struttura dell'edificio ma di fornire quegli scivoli/rampe senza i quali l'edificio sarebbe inaccessibile per diverse categorie di persone. Le rampe sono quindi i servizi web esposti (con l'integrato sistema di filtraggio) e il modo di usarle può essere diverso a seconda delle esigenze.

La figura rappresenta il "blueprint" dell'architettura. Di seguito sono discussi i diversi blocchi che la compongono.

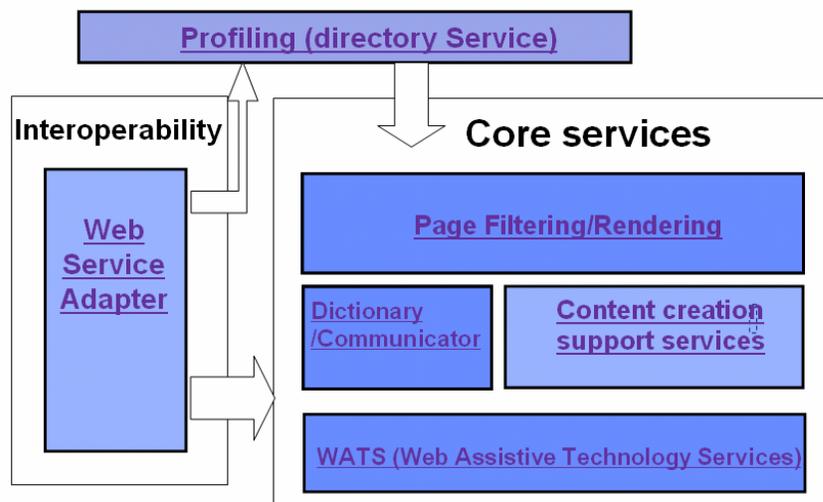


Figura 8: Schema dell'architettura di navigAble

- **Web Service Adapter:** l'insieme di tutti i servizi web le cui interfacce sono esposte al mondo esterno.
- **Profilazione:** servizio di registrazione dei profili utente e dei livelli di semplificazione associati. Si è pensato per la prima realizzazione di creare dei profili per servizio e non per utente in modo da non dover gestire dati sensibili (per es. nome, cognome). I tre set di attributi per profilo necessari inizialmente sono relativi ai profili: "standard", "semplificato", "molto semplificato". Il livello di complessità è evidentemente uno degli attributi del profilo. In futuro si potrebbe pensare di avere dei profili per utente che, invocando un opportuno servizio di supporto, possano essere mappati su profili per

servizio. È all'atto della profilazione che lato server è generato un cookie per l'identificazione univoca della tipologia d'utente; più tecnicamente, tutte le informazioni relative alla navigazione del profilo desiderato vengono generate lato server per essere salvate in un file lato client. Tutte le future interazioni client/server avranno così memoria del profilo nella comunicazione HTTP (il protocollo prevede invece che tutte le richieste di una pagina siano indipendenti).

- **Filtro sulle pagine (Page Filter):** sistema di filtraggio e personalizzazione delle pagine all'interno di uno stesso albero di navigazione. Il filtro interagisce con il sistema di profilazione in modo da ricavare il livello di complessità associato all'utente finale e invocare i servizi di interpretazione dei tag per il filtraggio dei contenuti e della struttura; questo permette di astrarre dalla scelta realizzativa del filtro e di centralizzare la logica alla base dell'interpretazione dei tag. Una volta definiti i contenuti, lato redazione, è il filtro che ne realizza la semplificazione per l'utente finale. Il Page Filter usa il Web Service Adapter, che a sua volta invoca i servizi core con chiamate SOAP. Il sistema di filtraggio integra in sé un **meccanismo di costruzione delle pagine** in base al target di utenza e una serie di **servizi di interpretazione dei tag**, mediante i quali possono essere interpretati tag proprietari per la personalizzazione delle strutture e dei contenuti delle pagine.
- **Dizionario:** è il dizionario multidimensionale grazie al quale ogni contenuto può essere associato a diversi formati (Testo, PCS, Bliss, Animazioni...). Il dizionario è da implementare come servizio core distribuito sulle diverse logiche, strati dell'architettura distribuita.

Il **blocco dell'interoperabilità**, interfaccia del sistema al mondo esterno, espone tutte le interfacce dei servizi invocabili dall'esterno come servizi basati su SOAP e realizza il meccanismo di Page Filter capace di garantire un adattamento dinamico delle pagine del sito al tipo d'utenza.

Il Page Filter, con l'ausilio di servizi interni, garantisce un *rendering ad hoc*, ovvero la strutturazione dinamica della pagina per l'utente riconosciuto in una prima fase di profilazione. Per la navigazione è garantita la presenza di alcuni profili standard dedicati agli utenti non ancora registrati.

Per ciascun utente, quindi, la pagina viene costruita "ad hoc", sia nella struttura che nei contenuti. La struttura è ricavata sfruttando i servizi di interpretazione dei tag, in grado, in base alla presenza di tag proprietari e alla profilazione dell'utente (standard o preregistrata), di definire pagine più o meno complesse.

Anche i contenuti, come la struttura, sono filtrati nella forma che più si addice alle caratteristiche dell'utente: questo grazie ai servizi del dizionario multidimensionale capaci di trasporre lo stesso contenuto informativo in più forme con l'ausilio di immagini reali o PCS, animazioni o Bliss.

L'architettura distribuita è concepita per poter supportare nel futuro i nuovi servizi che nel tempo concorreranno a coprire una più ampia gamma di funzionalità/applicazioni. È importante sottolineare la scalabilità della soluzione architettonica: nel caso di molti utenti concorrenti e di un grande numero di richieste è possibile, per mantenere alte le performance delle applicazioni, aumentare il numero di server a qualsiasi livello (web, application e DB), senza dover modificare i sorgenti.

2.4. Servizi per la Comunicazione Aumentativa Alternativa

navigAble è in grado di supportare diversi linguaggi di Comunicazione Aumentativa Alternativa (CAA), attraverso dei dizionari corredati da definizioni semplici ma efficaci che vengono associate a ciascun simbolo attraverso il legame funzionale fornito dall'ipertesto.

Ogni simbolo CAA che viene presentato a supporto della comprensione del testo scritto o del contenuto audio, infatti, può essere cliccato per poterne leggere o ascoltare la sua definizione.

I linguaggi CAA supportati da navigAble, a tendere, potranno essere:

- ◆ **PCS** (Picture Communication Symbols). È un set molto utilizzato perché di facile apprendimento e molto ricco di contenuti (sono quasi 4000 simboli). Molti significati sono doppi perché sono rappresentati sia da disegni dettagliati, che da sagome stilizzate. Per i simboli astratti, ovvero quelli che rappresentano l'azione, sono rappresentate una serie di situazioni concrete.
- ◆ **PICSYM** (Picture Symbols). Alcuni simboli sono costituiti da linee tratteggiate e linee intere per meglio rappresentare i rapporti topologici ; per esempio il simbolo che significa "dentro" è rappresentato da un quadrato tratteggiato contenente una sfera

nera. La linea tratteggiata è utilizzata anche per rappresentare una totalità di cui si vuole indicare solo una parte: la totalità è tratteggiata, la parte che si vuole indicare è invece disegnata con la linea intera (es. dormire)

- ◆ **BLISS**. I simboli sono una parte di un disegno stilizzato. Durante l'apprendimento è utile sovrapporre al simbolo la figura intera per comprendere la relazione con essa; in questo modo la rievocazione successiva viene facilitata. La combinazione di diversi simboli crea nuovi significati.
- ◆ **Immagini navigAbile**. Si tratta di una collezione di immagini fotografiche realizzata appositamente per navigAbile, che comprende le parole più comunemente utilizzate e che, a tendere, dovrebbe coprire l'intero dizionario di base della lingua italiana del DeMauro.

2.4.1. Semplificazione struttura e contenuti

Partendo dal presupposto che la disabilità crea modalità assai numerose di fruizione di un contenuto editoriale – sia esso un articolo, una lezione online, un libro elettronico o altro – il progetto navigAbile si è posto come obiettivo di arrivare a personalizzarne la complessità lessicale e del periodo, in modo da venire incontro alle necessità di utenti diversi.

Il servizio di semplificazione della struttura e dei contenuti editoriali, dunque, permette di presentare diverse versioni di uno stesso contenuto, con un livello progressivo di semplificazione, erogate sulla base del profilo di registrazione dell'utente.

La semplificazione dei contenuti presuppone **il lavoro di un team editoriale**, che per prima cosa si occupa di creare il contenuto in versione standard e quindi procede materialmente alla sua progressiva semplificazione, approntando per ciascun articolo una versione semplice e una molto semplice.

I livelli previsti sono:

- ◆ **Standard** – contenuto editoriale con un livello di complessità pari a quello utilizzato normalmente all'interno dei siti web informativi, dei portali, dei siti istituzionali, etc.
- ◆ **Easy** – contenuto editoriale semplificato, con un alto indice di leggibilità.
- ◆ **Basic** – contenuto estremamente semplificato, ridotto ai concetti base della frase.

Un **indice di leggibilità** è una formula matematica che attraverso un calcolo statistico è in grado di predire la reale difficoltà di un testo in base a una scala predefinita di valori. Per definire la formula di un indice di leggibilità si può tener conto di diverse variabili linguistiche,

cioè della misura di alcuni parametri del testo. Le variabili linguistiche più semplici sono, per esempio, lunghezza media delle parole e lunghezza media delle frasi. Vi sono variabili linguistiche che sono indipendenti dal contenuto del testo, come appunto le due citate, oppure variabili linguistiche legate al lessico, alla struttura del periodo, ecc. Sono state definite molte formule per la predizione della leggibilità, ma quelle che hanno avuto maggiore successo sono quelle che considerano variabili linguistiche di facile calcolo, come per esempio la lunghezza delle parole e delle frasi.

Per lo stesso motivo, i programmi che calcolano automaticamente il valore delle formule si limitano, nella maggior parte delle applicazioni, a formule semplici.

L'indice di leggibilità che è stato scelto per tale valutazione è **l'indice di Gulpease**.

2.4.2. Registrazione

In un sito web l'identificazione dell'interlocutore ha solitamente due obiettivi:

- **Riservatezza:** Permettere l'accesso a risorse personali e riservate all'interno del sito;
- **Personalizzazione:** Facilitare la fruizione di servizi e informazioni nelle modalità che meglio rispondono alle esigenze dell'interlocutore.

La modalità di registrazione attraverso la scelta di una **UserID e una password associate a un indirizzo e-mail**, divenuta ormai standard per il Web, sono state **semplificate** per tre motivi principali:

- per migliorarne **l'accessibilità**
- per evitare problemi legati alla normativa sulla **privacy** dei dati personali,
- perché nel contesto navigAbile, **non è sempre possibile associare un indirizzo e-mail** a un utente.

Si è dunque deciso di registrare ogni utente del sito navigAbile unicamente sulla base di un **nickname**, un soprannome, a cui viene associata una **password** scelta dall'utente. Tale accoppiata sarà ovviamente univoca.

Il servizio di registrazione comprende anche una **funzionalità di recupero password**, realizzata attraverso la modalità della "domanda personale", la cui risposta viene salvata al momento della registrazione. Questo sistema permette all'utente di recuperare la propria password senza aver bisogno di appoggiarsi a un indirizzo e-mail.

Qualcuno potrebbe criticare la validità e la sicurezza di tale soluzione: voglio sottolineare in proposito che nel profilo e nella parte "personale" del sito non vi sono dati sensibili (come

potrebbe accadere per un sistema di Home Banking), ma unicamente tutte quelle facilitazioni che un utente crea e aggiorna durante l'utilizzo di navigAbile (per esempio, la versione elettronica della propria tabella di comunicazione).

Per ogni utenza standard di ogni ragazzo, è prevista per gli operatori (educatori, logopedisti, etc.) la possibilità di autenticarsi come amministratore dell'utenza. Autenticandosi come amministratore di una determinata utenza, l'operatore ha accesso a tutte le funzionalità di personalizzazione dell'interfaccia e di gestione della tabella di comunicazione.

Alle funzionalità di amministrazione di una data utenza si accede utilizzando il soprannome dell'utente e aggiungendo alla fine di questo la parola "super" preceduta da un carattere *underscore* (trattino basso). La password resta la stessa dell'utente.

Esempio:

Soprannome: paolo	Amministratore: paolo_super
Password: topolino	Password: topolino
Accesso alle funzionalità standard associate al profilo di Paolo.	Accesso alle funzionalità di amministrazione e a quelle standard associate al profilo di Paolo

2.4.3. Personalizzazione

Il servizio permette di costruire **il profilo di accessibilità dell'utente** attraverso la raccolta di informazioni specifiche sulle sue abitudini di navigazione e di utilizzo del PC e sulle sue preferenze per quanto riguarda l'interfaccia in generale e la complessità dei contenuti.

La personalizzazione può essere effettuata in concomitanza con la registrazione dell'utente al sito oppure dopo un primo periodo di acclimatazione, quando l'utente si accorge di avere delle determinate esigenze che possono essere adeguatamente supportate da navigAbile.

Il servizio di Personalizzazione è inserito all'interno dell'indice dell'Area Servizi.

La prima schermata del servizio presenta l'elenco delle più comuni **modalità di navigazione** internet e dei dispositivi di input/output, inclusi tutti gli ausili più diffusi tra cui:

- Mouse o mouse modificato
- Tastiera o tastiera modificata
- Display Braille

-
- Caschetto e copritastiera metallico
 - Sensori di comando
 - Lettore vocale

Gli elementi di questo elenco possono essere selezionati a piacere, poiché non sono mutuamente esclusivi e, almeno per il momento, non provocano alcuna modifica del comportamento del sito.

La seconda pagina di configurazione permette di selezionare **il linguaggio di Comunicazione Aumentativa Alternativa (CAA)** utilizzato di preferenza.

Nella prima fase del progetto, il linguaggio supportato sarà soltanto il PCS, ma in un secondo momento è prevista l'integrazione di altri linguaggi quali il Bliss e il PICSYM.

Questa scelta si rifletterà direttamente nel tipo di linguaggio utilizzato per supportare la comprensione del testo all'interno del sito.

L'utente che non dovesse conoscere alcun linguaggio CAA oppure quello che legge e scrive abitualmente può selezionare l'opzione "nessun linguaggio".

La terza pagina di personalizzazione permette di scegliere **il tipo di contrasto** tra colore del testo e colore di sfondo della pagina che rende la lettura la più agevole possibile. Vengono proposti, nella prima fase del progetto, tre tipi di accoppiamento testo-sfondo, selezionati tra quelli la cui leggibilità è ritenuta massima (bianco o giallo su sfondo nero, nero su sfondo bianco). Il contrasto testo-sfondo, infatti, unitamente alle dimensioni del carattere, rappresenta spesso la chiave di volta dell'accessibilità a livello visivo del web.

Nella quarta pagina di configurazione può essere indicata la **dimensione preferita dei caratteri** utilizzati per il testo, scegliendo tra tre dimensioni predefinite: normale, grande e molto grande.

I tipi di caratteri scelti, **l'Arial e il Verdana**, non possono essere modificati. Questo perché l'Arial e soprattutto il Verdana, entrambi dei font senza grazie o "bastoni", proprio per questa loro caratteristica, consentono di raggiungere un'ottima leggibilità a monitor del testo.

La quinta pagina di personalizzazione permette all'utente di scegliere un **livello di complessità del testo** adatto alle proprie capacità di lettura e di apprendimento. Sono previsti tre livelli di complessità, selezionabili attraverso un radio button:

- **Normali** (Standard) – contenuto editoriale con un livello di complessità pari a quello utilizzato normalmente all'interno dei siti web informativi, dei portali, dei siti istituzionali, etc.

-
- **Semplificati** (Easy) – contenuto editoriale semplificato, con un alto indice di leggibilità.
 - **Molto semplificati** (Basic) – contenuto estremamente semplificato, ridotto ai concetti base della frase.

L'ultima pagina di personalizzazione permette all'utente di selezionare tutti gli **elementi aggiuntivi** che vuole avere in pagina a supporto della **fruizione del sito**. Nella prima versione di navigAbile l'utente potrà richiedere la presenza dei **simboli CAA** a supporto della lettura, dei **pulsanti audio** e del testo tutto **maiuscolo**.

Per facilitare l'introduzione del concetto di personalizzazione, nel sito pilota di navigAbile la procedura completa è stata sostituita **con una procedura semplificata**, basata sulla scelta di uno dei **tre profili standard** costruiti per rispondere alle esigenze della maggior parte degli utenti che hanno partecipato alla seconda fase dei test di usabilità.

Le **caratteristiche** dei tre profili standard sono le seguenti:

- **Profilo 1** – caratteri piccoli, contenuti editoriali livello Standard, immagini CAA assenti, supporto audio assente;
- **Profilo 2** – caratteri medio-grandi, contenuti editoriali livello Easy, immagini CAA assenti, supporto audio;
- **Profilo 3** – caratteri grandi, contenuti editoriali livello Basic, immagini navigAbile, supporto audio.

Data la complessità comunque associata alla scelta del profilo, **tale funzionalità resta comunque una prerogativa dell'educatore** o dell'operatore che assiste il ragazzo durante la fase di registrazione.

Importante sottolineare che la scelta in termini di personalizzazione non è mai definitiva: seguendo lo sviluppo delle capacità del ragazzo, l'operatore può sempre rientrare nella pagina riepilogativa Profilo, presentata all'interno della sezione Servizi, per modificare il profilo associato all'utente.

2.4.4. Autenticazione

L'autenticazione, cioè il momento in cui il sistema richiede l'inserimento del soprannome e della password, avviene nel momento in cui l'utente clicca per la prima volta sul link a navigAbile posto sul sito standard. Se l'utente non è ancora registrato, dalla pagina di

autenticazione potrà iniziare la procedura di registrazione al sito e, se vuole, anche quella di personalizzazione.

Se invece l'utente è già registrato dovrà inserire il suo soprannome e la password associata. Se l'autenticazione avviene con successo, il sistema navigAbile si riconfigura secondo i parametri di preferenza indicati dall'utente in fase di personalizzazione, sia dal punto di vista dei contenuti che della visualizzazione e della modalità di navigazione.

2.4.5. Dizionario visuale

La vera funzione di un dizionario non è quella di strumento di passiva e banale consultazione, ma bensì di supportare e arricchire la capacità espressiva degli utenti. In questo modo, il dizionario diventa un vero "il libro delle parole".

Il dizionario visuale propone l'intera collezione delle immagini navigAbile, che verrà costantemente ampliata fino a raggiungere le 2000 parole del dizionario di base della lingua italiana del DeMauro.

Per il futuro, inoltre, si prevede di avere un dizionario estensibile che possa essere personalizzato dall'utente per rispondere al meglio alle proprie esigenze.

Alcuni simboli più importanti o utilizzati più di frequente, infatti, potranno essere sostituiti con una foto (es: ciascun bambino potrà utilizzare al posto dell'immagine "mamma" la foto della propria mamma).

Oltre al dizionario visuale è prevista la realizzazione di un glossario che consenta l'approfondimento dei concetti chiave e delle "parole difficili" contenute all'interno di unità educative basate sul sistema navigAbile. Il linguaggio utilizzato in tali approfondimenti sarà consistente con quello utilizzato all'interno delle unità educative, che verrà tarato dai produttori di contenuti sulla base del percorso educativo e dalla capacità di apprendimento degli utenti a cui l'unità è rivolta.

2.4.6. La Tabella di Comunicazione

Uno degli ausili previsti dalla CAA è la tabella di comunicazione, un supporto sul quale inserire fotografie, disegni o simboli che possano aiutare la persona non parlante a comunicare bisogni, desideri, emozioni, esperienze. La tabella è un oggetto personale, non si può preconfezionare, nasce, cresce e si sviluppa con la persona che la utilizza. Nella tabella

devono essere inserite le parole che la persona vuole comunicare. Deve essere mobile e modificarsi a seconda delle esigenze: come il linguaggio è in continuo ampliamento, anche la tabella deve continuamente evolversi e arricchirsi.

In questo contesto si inserisce il servizio Tabella di Comunicazione di navigAbile, che si propone come una trasposizione elettronica della tabella di comunicazione cartacea utilizzata nella CAA.

All'interno di navigAbile, ogni utente può dotarsi di una tabella di comunicazione propria che contiene l'insieme di simboli o frasi conosciuti o usati più frequentemente. La tabella è personale e può essere costruita a partire dalle immagini del Dizionario Visuale, ed essere poi integrata con nuovi simboli che vengono incontrati, spiegati e appresi durante la navigazione sul sito.

La tabella contiene tutti i simboli conosciuti dal ragazzo, suddivisi per area (verbi, nomi, persone, oggetti, luoghi, etc.). In ciascuna area, i simboli sono presentati in ordine di inserimento, con in fondo quelli inseriti per ultimi nella Tabella.

Anche se la Tabella di Comunicazione nasce come base per i servizi del Compositore Frasi e della Bacheca, che si interfacciano con questa per permettere la selezione di un simbolo, esiste anche la possibilità di utilizzarla come strumento di comunicazione diretta e di apprendimento.

Nella modalità "**Indica**", infatti, il ragazzo può comunicare indicando i simboli visualizzati direttamente sul monitor oppure, se utilizza un dispositivo di puntamento (mouse o altro) o di scansione sequenziale (per esempio utilizzando il tabulatore e l'invio attraverso il caschetto e il copritastiera), può evidenziare i simboli che gli interessano, cliccando su di essi. I simboli cliccati rimarranno "accesi" in modo da formare una sequenza facilmente individuabile e comprensibile da parte dell'educatore o del logopedista.

Nella modalità "**Impara**", invece, cliccando su un simbolo l'utente ne visualizzerà la definizione inserita nel Dizionario Visuale e potrà decidere se, eventualmente, eliminarlo dalla Tabella.

Nella **modalità di amministrazione**, gli operatori possono creare le diverse aree in cui il ragazzo è abituato a tenere suddivisi i simboli che conosce. A ciascuna area della tabella di comunicazione può essere associato un nome e un simbolo che lo identifica in modo chiaro. I simboli nuovi verranno aggiunti attraverso una consultazione del dizionario e verranno inseriti

in fondo alla tabella. Analogamente esiste la possibilità per gli operatori di eliminare un simbolo dalla griglia.

Se una tabella è particolarmente ricca di simboli è previsto lo scrolling verticale della pagina. La tabella è formattata in modo da poter essere stampata, sia a colori che in bianco e nero, in modo semplice e ben organizzato.

Alcune funzionalità aggiuntive, quali la possibilità di organizzare i simboli in maniera ottimale o di lasciare delle caselle vuote per separare gruppi di simboli all'interno di un'area della tabella, sono attualmente in fase di studio.

2.4.7. Compositore di frasi

Questo servizio permette di comporre una frase, sia come testo alfabetico sia a partire dai simboli presenti sulla tabella di comunicazione. In fase di personalizzazione, infatti, è possibile indicare se l'utente vuole utilizzare il Compositore simbolico o quello testuale: se questa preferenza non viene specificata all'inizio, la scelta verrà presentata al momento in cui l'utente esprime la volontà di scrivere una frase.

Le frasi composte possono essere salvate in un elenco personale di frasi (testuali o simboliche) utilizzate di frequente, e possono essere facilmente cancellate all'esigenza.

Se in fase di personalizzazione è stato scelto l'utilizzo del supporto audio, ogni frase salvata in tale elenco può essere letta "a voce alta" dal sintetizzatore vocale.

In una successiva fase di sviluppo, per venire incontro alle necessità di utenti come i dislessici, il Compositore di frasi testuale potrebbe essere dotato di un sistema di predizione delle parole.

2.4.8. Bacheca

La Bacheca è l'equivalente virtuale di un tabellone esposto in un luogo pubblico, dove tutti possono esporre una propria opera (un disegno, una fotografia, una poesia) in modo che gli altri la possano guardare ed eventualmente commentare.

Il servizio di Bacheca di navigAble quindi prevede la possibilità di pubblicare immagini (disegni, fotografie, etc.) e documenti (temi, racconti, poesie, etc.) in modo che risultino visibili sul Web e possano essere fruiti da tutti gli utenti che appartengono alle diverse comunità che utilizzano navigAble.

I testi pubblicati vengono scritti attraverso il compositore di frasi, sia sotto forma di testo alfabetico che di frasi costruite con i simboli della tabella di comunicazione elettronica.

Ciascun utente ha la possibilità di inserire un oggetto ex-novo nel Forum e di commentare tutti gli oggetti inseriti dagli altri utenti. I commenti vengono riportati in ordine cronologico, con il più recente per primo.

2.4.9. Forum

La realizzazione del servizio di Forum è prevista per la seconda fase di sviluppo di navigAbile. Un Forum, o gruppo di discussione, è un luogo d'incontro virtuale dove un certo numero di persone si ritrovano per discutere problemi d'interesse comune. In pratica, si tratta di un'area web, solitamente afferente a un certo sito, in cui ciascun utente può inserire un messaggio che potrà essere letto da tutti gli altri utenti che eventualmente commenteranno a loro volta, dando vita a discussioni di ogni genere e lunghezza sugli argomenti più disparati.

2.4.10. E-mail

La realizzazione del servizio di E-mail è prevista per la seconda fase di sviluppo di navigAbile. Il servizio E-mail di navigAbile permette di inviare una frase costruita attraverso il Compositore Frasi direttamente a un indirizzo di posta elettronica.

2.5. Il prototipo statico iniziale

Le due prime sessioni di test di usabilità di navigAbile si sono svolte su una **versione prototipale statica del sito web**, i cui contenuti sono stati realizzati ad hoc con il supporto degli esperti (per quanto riguarda il Don Gnocchi) e dei professori (per i contenuti dell'Istituto Matilde di Canossa di Reggio Emilia).

Lo **sviluppo dell'albero di navigazione** del prototipo viene presentato in figura: in esso si evidenzia la suddivisione dei contenuti in quattro aree fondamentali:

- ◆ **Servizi:** contiene soltanto una versione statica del servizio **Bacheca**, che mostra quattro disegni realizzati da ragazzi disabili e caricati sul sito. Per ogni disegno è inserita una breve descrizione audio che lo illustra.

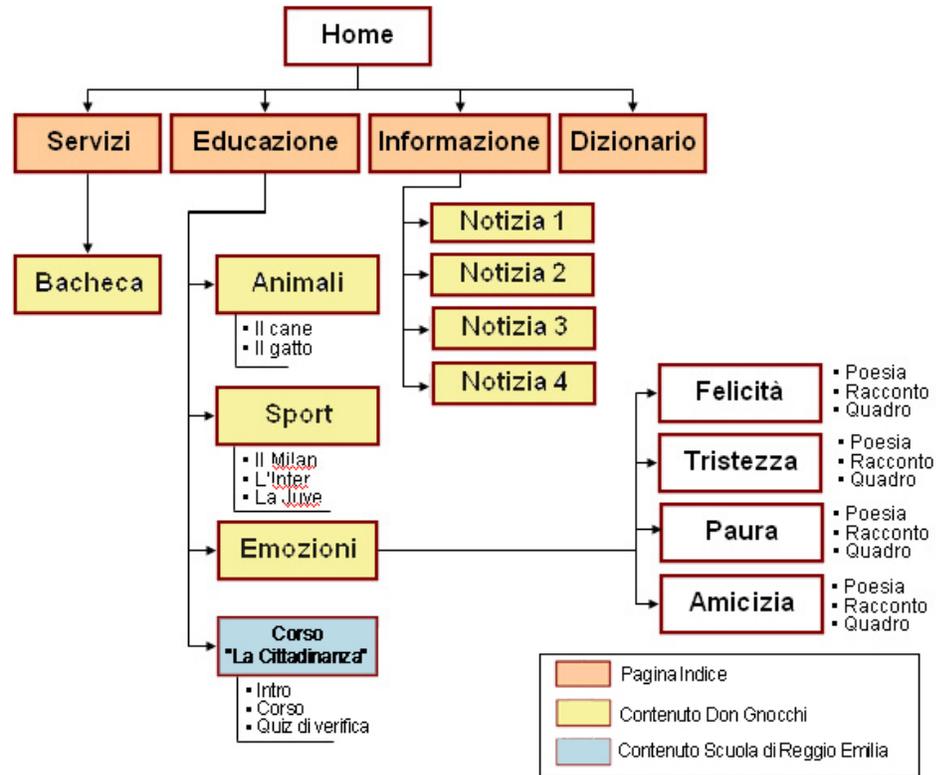


Figura 9: Sviluppo dell'albero di navigazione del prototipo iniziale

- ◆ **Educazione:** contiene un'ampia area di contenuti educativi sviluppati per i ragazzi dell'Istituto Don Gnocchi. Quest'area è a sua volta suddivisa in tre categorie: Animali (dove sono inserite due unità educative dedicate ai cani e ai gatti), Sport (dove sono inserite delle schede storiche sulle squadre di calcio più gettonate), Emozioni (suddivisa a sua volta in quattro sottosezioni, una per ogni emozione presentata). Oltre ai contenuti realizzati per il Don Gnocchi, in quest'area è presente anche la prima sessione del corso di storia "La Cittadinanza", realizzato in collaborazione con i professori dell'Istituto Matilde di Canossa di Reggio Emilia per permettere ai loro ragazzi di provare navigAbile.
- ◆ **Informazione:** contiene quattro notizie curiose o di attualità.
- ◆ **Dizionario:** contiene una selezione di 500 PCS della Mayer-Jonhson, organizzati secondo le categorie standard di utilizzo.



Figura 10: Homepage del prototipo in versione semplificata

L'interfaccia del prototipo è basata su **due tipologie di pagina**: la pagina Indice e la pagina Contenuto.

La **pagina Indice** presenta i link ai contenuti sotto forma di **tasti**: il contenuto a cui il tasto rimanda è indicato sia con una parola sia con il suo corrispondente PCS. A ogni tasto contenuto nella pagina è associato un pulsante **audio**, la cui funzione è leggere il nome del link. In questa versione prototipale l'audio associato ai diversi tasti è stato **pre-registrato in formato WAV**.

Per suggerire in modo inequivocabile l'interattività a essi associata, i tasti sono dotati di un **effetto roll-over**: quando vi si passa sopra con il mouse o li si seleziona con il tabulatore, intorno a essi si accende un **bordo rosso** ben visibile, che ovviamente scompare quando il mouse lascia l'area attiva del tasto (e quindi del link).

Quando un indice presenta **un numero di link maggiore di quattro**, i tasti in eccedenza vengono sistemati in una 'pagina successiva' a cui si accede attraverso il link posto sul **tasto "Avanti"**. Questo tasto compare a destra nella barra di navigazione in testata, in posizione speculare rispetto al tasto "Indietro".



Figura 11: Una pagina di indice del prototipo

Cliccando su uno dei tasti dell'Indice si accede a **una pagina Contenuto**.

L'interfaccia di navigazione della pagina Contenuto è la stessa della pagina Indice, con i tasti "Home", "Indietro" e "Avanti" collocati nella testata. Diversa è però la disposizione degli oggetti in pagina, che in questo caso non sono più tasti di navigazione, ma brani di testo e immagini.

Ogni pagina Contenuto è aperta da **una fotografia** e dal **titolo del contenuto** presentato in pagina. A fianco del titolo è presente il **pulsante audio** che, cliccato, attiva la lettura di tutta la pagina. Il supporto audio è realizzato attraverso dei **file WAV preregistrati da uno speaker**.



Figura 12: Una pagina di contenuto del prototipo

Seguono le **immagini PCS che illustrano le parole chiave** del testo, supportando il ragazzo disabile nella fruizione del contenuto, e immediatamente di seguito vi è il **testo vero e proprio**, proposto con un linguaggio semplificato, riportato con caratteri grandi e di un colore che crea un buon contrasto con lo sfondo bianco. Le **parole chiave** del testo, che corrispondono alle immagini PCS, vengono indicate in rosso.

Per consentire ai ragazzi di apprendere nuove parole, i PCS relativi alle parole chiave più difficili contengono un link diretto al Dizionario, dove ne viene presentata una semplice definizione. La presenza di tale link è indicata dal fatto che i **PCS collegati al Dizionario** sono dei tasti, con il relativo effetto roll-over sul bordo. Le pagine di Dizionario relative ai contenuti del prototipo sono state realizzate in modalità statica, secondo il modello di pagina sviluppato appositamente per questo servizio.



Figura 13: Esempio di una voce del dizionario di navigAbile

In questo modello di pagina, oltre al titolo della sezione **“Dizionario”** corredato da un’icona, viene riportata la parola difficile e una sua **definizione semplificata**. Anche in questo caso, il pulsante audio attiva la lettura pre-registrata del contenuto della pagina.

2.6. Il sito pilota

Sulla base dei risultati dei test di usabilità e accessibilità sul prototipo iniziale del sito navigAbile, è stato realizzato il sito pilota, comprendente tutte le funzionalità e i servizi previsti per il rilascio della prima versione on-line (il portale è raggiungibile all'indirizzo www.navigabile.it).

Lo sviluppo dell'albero di navigazione del sito pilota viene presentato in figura.

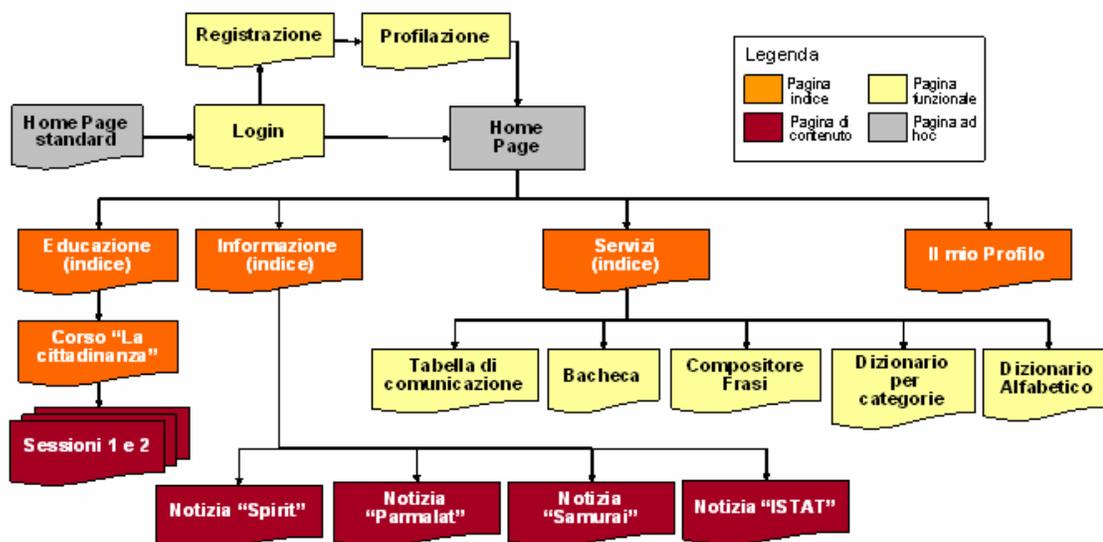


Figura 14: Albero di navigazione del sito pilota navigabile.it

Come nel prototipo, la Home Page presenta i contenuti del sito suddivisi in quattro aree:

- **Informazione:** contiene notizie curiose o di attualità su cui è applicato il servizio di *Page filtering*;
- **Servizi:** contiene la Tabella di Comunicazione, la Bacheca, il Compositore Frasi e il Dizionario Visuale nelle sue due modalità di consultazione, alfabetica e per categorie.
- **Educazione:** Contiene le quattro sessioni del corso di storia "La Cittadinanza", realizzato in collaborazione con i professori dell'Istituto Matilde di Canossa di Reggio Emilia per permettere ai loro ragazzi di testare navigAbile.
- **Profilo:** contiene una pagina di riepilogo delle informazioni di personalizzazione e permette di accedere alla pagina di selezione del profilo tra i tre profili standard creati appositamente per il sito pilota.

Possiamo notare come il Dizionario sia stato collocato all'interno dell'Area Servizi.

Osservando l'albero di navigazione del sito pilota possiamo notare la presenza delle funzionalità di gestione utenti: la **registrazione utenti**, la **profilazione** e il **login** (autenticazione al sito), che permettono di creare e di accedere a una versione 'profilata' del sito in base alle esigenze dell'utente.

La **home page standard**, quella a cui si accede digitando l'indirizzo web del sito pilota, si presenta con una grafica e un layout in linea con le tendenze attuali del web.



Figura 15: Homepage di navigAbile in versione standard

Un tasto ben visibile e di grandi dimensioni ("Entra nella versione navigAbile"), permette di accedere alla funzionalità di autenticazione. Esso è inserito ad hoc in una posizione immediatamente raggiungibile per gli utenti che navigano su Internet utilizzando un sistema di scansione sequenziale dei link. Dalla pagina di autenticazione si accede alla **procedura di registrazione**, in cui un nuovo utente può scegliere un nome utente e una password per crearsi la propria utenza di navigAbile. Se un utente è già registrato, da questa pagina può autenticarsi in modo da visualizzare il sito, i contenuti e la navigazione secondo i propri parametri di personalizzazione, e accedere ai propri servizi 'personali' quali la Tabella di Comunicazione e l'Elenco Frasi.

La procedura di registrazione si conclude con un invito alla **personalizzazione del sito**.



Figura 16: Pagina per la scelta del profilo

Per facilitare l'introduzione del concetto di personalizzazione, nel sito pilota di navigAbile si è realizzata una **procedura semplificata**, basata sulla scelta di uno dei **tre profili standard** costruiti per rispondere alle esigenze della maggior parte degli utenti che hanno partecipato alla seconda fase dei test di usabilità.

Le **caratteristiche** dei tre profili standard sono le seguenti:

- **Profilo 1:** caratteri piccoli, contenuti editoriali livello Standard, immagini CAA assenti, supporto audio assente;
- **Profilo 2:** caratteri medio-grandi, contenuti editoriali livello Easy, immagini CAA assenti, supporto audio;
- **Profilo 3:** caratteri grandi, contenuti editoriali livello Basic, immagini navigAbile, supporto audio.

Data la complessità comunque associata alla scelta del profilo, l'utilizzo di **tale funzionalità resta comunque una prerogativa dell'educatore** o dell'operatore che assiste il ragazzo durante la fase di registrazione.

Il supporto audio, nel sito pilota, è realizzato attraverso ScanSoft Realspeak, un software di sintesi vocale residente sul server di navigAbile. Questo software costruisce in tempo reale i file audio associati al contenuto (in formato WAV) che poi vengono inviati in *streaming* al computer dell'utente e riprodotti dal programma Windows Media Player direttamente nella finestra di Internet Explorer che visualizza il contenuto stesso.

Alla fine della prima personalizzazione, o dopo l'autenticazione se l'utente è già registrato, viene visualizzata la home page in versione navigAbile. Nella figura che segue vediamo le due diverse visualizzazioni della home page corrispondenti al profilo 3 (con le immagini) e al profilo 2 (senza immagini). La home page corrispondente al profilo 1 è del tutto identica a quella pubblica, a parte le funzionalità legate all'utente (gestione profilo e logout).



Figura 17: Homepage di navigAbile utilizzando il profilo 2



Figura 18: Homepage di navigAbile utilizzando il profilo 3

Cliccando sul tasto "**Informazione**" si accede all'area del sito pilota che contiene le notizie di attualità. Nell'area Informazione sono contenute quattro notizie importanti o curiose, selezionate opportunamente in modo da rivestire un interesse per i ragazzi.

Il **sistema di Filtro sulle pagine** descritto nel capitolo dedicato all'architettura di navigAbile permette di visualizzare una versione diversa dei contenuti a seconda del profilo impostato dall'utente.

In figura vediamo la **versione standard** del testo (Normal) che corrisponde al **profilo 1**, la **versione sintetica** per il **profilo 2** (Easy) e la **versione sintetica semplificata** per il **profilo 3** (Basic).

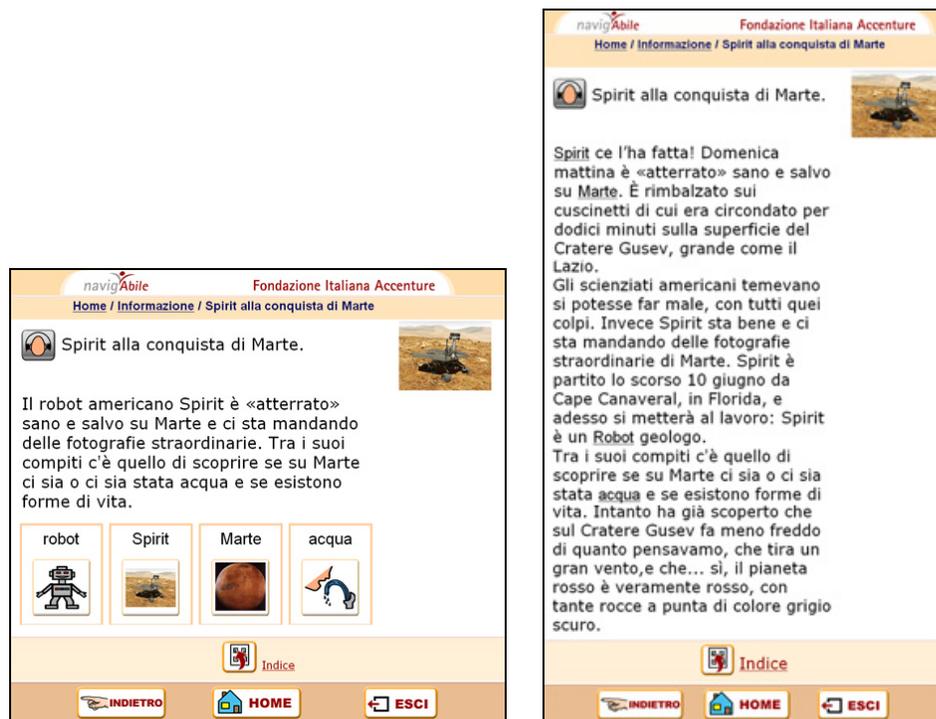


Figura 19: Pagina di una news in versione Normal (profilo 1)

Cliccando sul tasto “**Servizi**” posto nella Home Page si accede alla pagina **Indice Servizi**, che contiene il punto d’accesso a tutti i servizi di navigAbile inseriti nel sito pilota.



Figura 20: Indice dei servizi di navigAbile (con profilo 2)

Questi servizi sono:

- ◆ **Compositore Frasi:** contiene la funzionalità completa di scrittura frasi, sia testuali che simboliche, utilizzando le immagini navigAbile presenti sulla Tabella di Comunicazione.
- ◆ **Bacheca:** permette agli utenti di caricare un file di testo in formato DOC, XLS, PPT, PDF e TXT e immagini in formato GIF e JPG. Permette inoltre di inserire commenti a ciascun file caricato e di leggere i commenti lasciati dagli altri utenti.
- ◆ **Tabella di Comunicazione:** contiene la tabella di comunicazione dell'utente, da popolare attraverso il dizionario con le immagini navigAbile corrispondenti ai simboli CAA presenti sulla tabella di comunicazione cartacea dell'utente.
- ◆ **Dizionario per Categorie:** contiene una selezione di immagini fotografiche selezionate appositamente per navigAbile e organizzate secondo le categorie standard di utilizzo della CAA.
- ◆ **Dizionario alfabetico:** contiene una selezione di immagini fotografiche selezionate appositamente per navigAbile organizzate in ordine alfabetico per una più facile consultazione da parte dell'operatore che affianca l'utente disabile nella navigazione.

La **personalizzazione dell'interfaccia** agisce sulle pagine di tipo Indice come la Home Page (navigazione) e sulle pagine di tipo Contenuto come le notizie inserite nell'area Informazione (filtro sul testo), ma non sull'area Servizi: questo significa che, all'interno del sito

pilota, sia l'Indice Servizi sia tutti i servizi che esso contiene non saranno modificati dalla personalizzazione.

La sezione **Educazione**, contiene il corso di storia **"La Cittadinanza"**, articolato in quattro sessioni e realizzato in collaborazione con i professori dell'Istituto Magistrale Matilde di Canossa di Reggio Emilia, per permettere ai loro ragazzi di provare navigAbile.

Anche sul corso è applicato il sistema di Filtro, che agisce unicamente sul layout dell'interfaccia.



Figura 21: Pagina di indice delle sessioni del corso

I contenuti, invece, **restano invariati** per necessità di consistenza didattica.

Ciascuna delle quattro sessioni comprende una **pagina indice** che presenta l'elenco dei capitoli della sessione: è possibile entrare in un capitolo cliccando sull'immagine associata o direttamente sul link inserito sul titolo.

A partire dall'Indice sono previste **due modalità di fruizione** del corso:

- ◆ **Modalità puntuale** – il corso viene fruito capitolo per capitolo, anche senza un ordine preciso, entrando in un capitolo attraverso l'indice e ritornando all'indice per entrare in quello successivo.
- ◆ **Modalità sequenziale** – il corso viene fruito in modo ordinato e sequenziale, scorrendo le pagine dei capitoli a una a una, spostandosi in avanti (o eventualmente indietro) attraverso le frecce poste nella barra di navigazione di ogni pagina.

Le pagine di contenuto includono il **supporto audio** e sono corredate da **immagini e grafici** che accompagnano il testo per sottolineare o illustrare i concetti più importanti.



Figura 22: Una pagina di contenuto del corso

A tutte le parole difficili o che necessitano un approfondimento è associata una pagina di glossario, accessibile attraverso un link posto sulla parola stessa. Le parole collegate al glossario sono riconoscibili dalla presenza della sottolineatura.

Al termine di ogni sessione è previsto un **quiz di verifica**: per poter concludere con successo la sessione è indispensabile superare il test a pieni voti.

2.7. Gestione del progetto: le specificità di navigAbile

Fondamentalmente si possono individuare tre componenti fondamentali che hanno caratterizzato il project management di navigAbile:

- ◆ interdisciplinerietà
- ◆ specificità del testing: valutazione dell'esperienza della coppia utente-tutor
- ◆ validazione scientifica sulla base dei risultati del testing

La prima peculiarità del progetto navigAbile è la distinzione e la complementarietà di competenze tra i componenti del team di progetto. In particolare il team della Fondazione Italiana Accenture ha messo in campo le competenze tecnologiche e di usabilità e accessibilità

web, intese come tecniche e regole di progettazione per massimizzare l'usabilità e l'accessibilità di un sito web. Il team messo in campo invece dagli altri partner di progetto, Fondazione Don Gnocchi e Fondazione Francesca Rava, ha invece portato le competenze e le esperienze concrete derivanti dalla conoscenza e dal lavoro quotidiano con il gruppo di utenti che sono stati coinvolti in questa fase di progetto, ossia i ragazzi e gli operatori del Don Gnocchi e i ragazzi e i professori delle scuole coinvolte. Questa distinzione di ruoli è stata una condizione inevitabile di cui si è preso atto fin dall'inizio del progetto, in quanto le persone della Fondazione Italiana Accenture, che sono professionisti che vengono dal mondo della consulenza e che hanno quindi elevate competenze nella realizzazione di progetti web, non possedevano evidentemente competenze specifiche relative al mondo della disabilità, né volevano improvvisarsi "esperti" di un settore così complesso. D'altro canto, i professionisti coinvolti dai partner di progetto, pur non essendo "esperti di web", avevano indiscutibili competenze ed esperienze riguardo agli aspetti sia teorici che clinico-terapeutici delle esigenze del gruppo di utenti coinvolti. Questa distinzione di ruoli e competenze, situazione che in genere non si verifica nei progetti "normali" dove l'analista e il progettista hanno comunque anche delle conoscenze della realtà del cliente, è stata vista da una parte come un punto di attenzione nella gestione del progetto, dall'altra come un'opportunità e una garanzia di validità del lavoro svolto.

La seconda specificità riguarda la fase di test. La progettazione, l'esecuzione e l'analisi dei risultati delle sessioni di test è stata oggetto di un attento studio per adattare, e spesso rivedere in modo sostanziale, la metodologia utilizzata per effettuare test di usabilità su applicazioni *web-based*. In primo luogo, non si parla di un singolo utente coinvolto nel test ma di una coppia ragazzo-operatore. In secondo luogo durante il test, oltre alle osservazioni e alle considerazioni relative all'usabilità del prodotto, sono state raccolte anche osservazioni di diverso tipo (psicopedagogiche), da qui la necessità di avere due osservatori presenti con ruoli diversi. La valutazione dei risultati ha poi tenuto conto di criteri qualitativi e non solo quantitativi.

La terza specificità del progetto, legata alle prime due, consiste nell'introduzione di un passo di validazione scientifica dopo la fase di test. La validazione scientifica può essere strutturata in modo molto diverso: si va dall'analisi delle osservazioni effettuate durante il test e dei commenti raccolti, alle sessioni di brainstorming con gli operatori, a un periodo anche prolungato di utilizzo del sistema al fine di verificarne sul campo la reale efficacia e

rispondenza ai requisiti degli utenti. La scelta delle diverse strategie di validazione dipende, oltre che dalla criticità dei requisiti affrontati, anche dall'importanza del rilascio (rilascio intermedio, rilascio finale). La validazione scientifica presenta delle similitudini con l'utilizzo di rilasci pilota resi disponibili a un gruppo ristretto di utenti, pratica comune nella gestione dei progetti. Come nel caso dei rilasci pilota, è previsto naturalmente un ciclo di revisione del prodotto in modo da raccogliere le evidenze emerse durante la validazione.

2.8. Sviluppi futuri di navigAbile

Guardando a quelli che sono le tendenze e i temi caldi nel campo dell'accessibilità web, dell'ICT (Information and Communication Technology) e in base all'esperienza maturata in questa prima fase di progetto con i partner e con i ragazzi che hanno partecipato, sono ipotizzabili diverse direttrici di evoluzione di navigabile, raggruppabili in quattro aree:

- **Diffusione dei servizi navigAbile ad altri siti e portali:** i servizi architetturali di navigAbile sono stati progettati e realizzati fin dall'inizio considerando che dovessero essere fruibili sia dai primi utenti che hanno collaborato con noi, costituendo per loro una soluzione *end-to-end* a un insieme di bisogni identificati insieme ai ragazzi e agli operatori, sia da altri siti o portali che decideranno di rendersi accessibili. Il tema, data l'importanza, è trattato in modo più approfondito nel paragrafo seguente: "Strategie di diffusione";
- **Identificazione di nuovi cluster di utenti:** la prima fase di progetto si è focalizzata su un cluster di utenti, suddiviso ulteriormente in diversi sottocluster, con problemi di comunicazione e relazione accomunati dal fatto di utilizzare strumenti di comunicazione multimodale. Le fasi successive di progetto potrebbero indirizzare in modo specifico i bisogni di altri gruppi di utenti, in parte riutilizzando i servizi già sviluppati e in parte sviluppando servizi nuovi. Alcuni esempi di cluster interessanti sono utenti affetti da problemi di dislessia, di afasia, di ipovisione. Alcuni dei servizi elencati nel punto "Evoluzione dei servizi" possono essere ricondotti a uno o più di questi cluster. L'applicazione del metodo navigAbile a questi nuovi utenti prevede naturalmente una fase di analisi preliminare dei bisogni e di identificazione dei requisiti specifici;

-
- **Evoluzione del sito pilota in un portale per l'accesso a Internet:** il sito pilota potrebbe diventare un vero e proprio punto di partenza da cui accedere a Internet, fornendo una serie di servizi quali:
 - Servizi di community, quali e-mail e forum, accessibili e multimodali;
 - Sezione "giochi", con giochi fruibili in versione web o scaricabili;
 - Gestione dei "Preferiti", in modo da poter garantire al navigatore con bisogni speciali la possibilità di accedere in modo semplice (minimizzando la digitazione) ai siti web preferiti;
 - Barra di navigAbile aggiunta al browser (tipo barra di Google) o posta in testa ai siti in navigazione, per accedere in modo comodo ad alcuni servizi di navigAbile, quale per esempio il dizionario multimodale per accedere alle definizioni (testuale e per immagini) dei termini;
 - Servizio di directory dei siti 'navigAbili' (ossia che utilizzano l'architettura di navigAbile)
 - Motore di ricerca navigAbile;
 - **Evoluzione dei Servizi architetturali:**
 - Evoluzione dei servizi attuali: alcune delle evoluzioni possibili sono:
 - Tabella di comunicazione: potenziamento delle possibilità di personalizzazione (in particolare possibilità di caricare immagini personali);
 - Servizio di e-mail: possibilità di inviare file in allegato – possibilità di ricevere e leggere mail;
 - *Repository* utenti: il *repository* utenti di navigAbile diventerà un servizio di directory offerto a tutti i siti navigAbili per accedere, in modo protetto e garantendo la privacy, al profilo di accessibilità degli utenti registrati;
 - Servizi di supporto all'utilizzo del *tag system* per il *filtering* delle pagine (editor, strumenti di anteprima, controllo errori...);
 - Nuovi servizi:
 - Gestione della multicanalità: l'utilizzo di dispositivi alternativi, quali Tablet PC, Pocket PC, telefonini UMTS, web-television ecc. apre certamente delle interessanti prospettive per aumentare l'accessibilità dei servizi di navigAbile;

-
- Gestione SMS e MMS: oltre alla gestione delle mail, i servizi di interazione possono essere estesi per gestire SMS e MMS;
 - Multilingue: attualmente l'architettura navigAble è progettata per essere estensibili ad altri paesi/lingue, ma alcuni componenti (es. il dizionario multimodale) devono essere localizzati. Inoltre il test e la sperimentazione in un altro contesto linguistico e culturale deve essere effettuato con un partner locale;
 - Predizione delle parole: la predizione delle parole è un servizio importante per un ampio spettro di disabilità.
 - Servizi di riconoscimento del parlato/dettatura/navigazione vocale: sono servizi interessanti per aumentare le possibilità di alcuni particolari gruppi di utenti;
 - CSS Manager: la gestione dei fogli di stile (CSS) può aumentare di molto l'accessibilità di una pagina web, giocando su fattori di presentazione (colori, dimensione font, ecc.) che sono spesso determinanti per molti utenti.

2.9. Attività a cui ho partecipato e componenti architetture che ho sviluppato

Nell'ambito dello stage svolto in collaborazione con Accenture ho personalmente partecipato alle seguenti attività:

- **Analisi dei requisiti:** si tratta di uno studio che precede ogni altra attività del progetto e che è volto a mettere in luce quali sono le esigenze e i fabbisogni che un disabile ha nella normale navigazione. Questa attività si è concretizzata con una documentazione iniziale sul tema dell'accessibilità (norme, validatori, etc...) e una valutazione dello stato di maturità dei portali italiani ed esteri in tema di accessibilità, con particolare attenzione a quelli degli enti pubblici. Questo studio di fattibilità ha avuto come output una serie di requisiti, che sono stati poi condivisi ed integrati con gli esperti della Fondazione Don Gnocchi. I requisiti rivisti sono stato l'input per la fase di progettazione del prototipo iniziale.
- **Dizionario multimodale:** ho ricavato da web, con un wrapper realizzato ad hoc, la base di dati lessicale per l'italiano. In particolare ho ricavato le tabelle: FORMA_FLESSA, LEMMA, SINONIMI, DESCRIZIONI. Ho progettato lo schema logico della base di dati di navigAbile, per la parte relativa al Dizionario Multimodale. Con riferimento alla figura che rappresenta lo schema dell'architettura dei servizi, ho realizzato il componente "Dictionary/Communicator".
- **Servizio di content management per il caricamento batch delle immagini nel dizionario:** si tratta di un'utility che aggiorna il database di navigAbile con le immagini categorizzate che si trovano in una certa cartella. All'interno di questa cartella devono essere presenti delle sottocartelle con lo stesso nome delle categorie del dizionario. Tutte le foto che vengono trovate nelle sottocartelle vengono associate al lemma corrispondente: questo consiste nell'aggiornare il campo URL_PCS del lemma (contenente il riferimento all'immagine PCS associata al lemma). Vengono poi aggiornati i campi che specificano la categoria del dizionario a cui appartiene il lemma.
- **Sviluppo del componente per l'analisi del testo.** Si tratta di un tool redazionale che calcola l'indice di leggibilità e quello di comprensibilità di un certo testo. Si tratta di un'interfaccia web-based con un textbox in cui viene inserito il testo; cliccando su un pulsante è possibile visualizzare le statistiche sugli indici di navigAbile.

-
- Sviluppo dei wrapper per ricavare la base di dati linguistica: ho realizzato tre diversi programmi per popolare le tabelle FORMA_FLESSA, LEMMA_ORIGINE, SINONIMI, DESCRIZIONI, ricavandole da tre diverse banche dati disponibili online.

Inoltre ho personalmente sviluppato l'intero studio sulla metodologia di analisi di complessità e leggibilità dei testi. Su queste tematiche si stanno concentrando sforzi e risorse per gli sviluppi futuri del progetto navigAbile.

Con riferimento alla figura che rappresenta l'Architettura dei servizi di navigAbile, i componenti di supporto alla gestione dei contenuti, si collocano all'interno del componente "Content Creation support service".

Tutto il codice sviluppato è stato scritto con il linguaggio C#; l'ambiente di sviluppo impiegato è Microsoft Visual Studio .NET 2003.

3. Il contesto

3.1. La Comunicazione Aumentativa Alternativa (CAA)

La Comunicazione Aumentativa Alternativa, come dice la parola stessa, è un insieme di metodi, strategie e ausili di comunicazione che servono per potenziare le capacità del soggetto di comunicare e offre anche un metodo alternativo al linguaggio dove esso sia assente e molto improbabile da verificarsi.

Da parecchi anni ormai la disabilità verbale viene affrontata come un problema di comunicazione, e nei casi in cui non è possibile ripristinare o far evolvere la funzione verbale, mediante terapia riabilitativa, si sono cercate altre soluzioni che consentissero di ricreare un ponte tra il pensiero del disabile e il modo esterno.

Quando una persona non può accedere ai normali mezzi espressivi (parola - scrittura) e si desidera comunicare può ricorrere ad un codice alternativo per rendere trasmissibile il proprio pensiero.

Le esperienze hanno però dimostrato che se è possibile sostituire il linguaggio verbale con altre modalità (gestuali, grafiche, simboliche) è necessario che questo avvenga attraverso un processo di integrazione che si articola a più livelli: funzionale, emotivo e cognitivo. Vanno perciò tenuti in considerazione tutta una serie di fattori, relativi all'individuo e al suo ambiente, per poter modulare tale integrazione ed evitare il fallimento di un lavoro mirato a un recupero comunicativo.

Nella trattazione che segue si farà riferimento in particolare alle tematiche legate all'utilizzo della CAA da parte di disabili in età evolutiva, poiché si tratta del cluster di utenti a cui è principalmente rivolto navigAbile.

L'uso di simboli, fotografie, gesti, apparecchi informatici serve a fare in modo che il bambino sperimenti un modo di comunicare comprensibile a tutti, così da non essere sempre dipendente dai genitori e familiari che ogni volta devono tradurre per altre persone i suoi desideri o i suoi pensieri.

Usare simboli o fotografie non preclude la comparsa o il rafforzamento del linguaggio, infatti i bambini che sperimentano quanto può essere utile ed efficace dire qualcosa a qualcuno cercheranno di aumentare la loro capacità comunicativa il più possibile.

Può essere giusto ricorrere alla CAA quando un bambino non riesce a sviluppare il linguaggio verbale o quando esso non sia sufficiente a permettergli la comunicazione con gli altri, sia perchè povero di vocaboli, sia perchè incomprensibile per chi non lo frequenta abitualmente. Specialmente nella socialità con i coetanei la comunicazione diventa uno strumento insostituibile, se esso è assente il bambino non potrà entrare in relazione con gli altri bambini e quindi crescere all'interno dell'ambiente sociale. Un bambino che non parla viene molte volte giudicato come un bambino che non capisce, mentre ciò non è vero in generale.

I sistemi iconici, in sostituzione del sistema alfabetico, strutturalmente molto più complesso, per la loro facilità di comprensione del significato rappresentato, costituiscono una notevole facilitazione nell'ambito della CAA.

In una società matura, le difficoltà a comunicare non possono rappresentare una barriera alle opportunità e alla realizzazione di ciascuno di noi. La Comunicazione Aumentativa Alternativa può essere vista come una "funzione" che garantisce l'accesso della persona con disabilità comunicativa alle sue competenze cognitive, affettive, pragmatiche peraltro indispensabili per ricostruire una nuova identità personale e sociale.

3.2. Linguaggi iconici: il caso di PCS e BLISS

Diversi linguaggi possono consentire il ripristino della funzione comunicativa: essi possono essere, mediati direttamente dal corpo (gestuali), o rappresentati da segni grafici (disegni, figure, simboli) per essere indicati.

Uno dei sistemi di comunicazione alternativa oggi maggiormente utilizzato nei casi di disabilità è il linguaggio iconico Bliss (inizialmente denominati Semantography dal suo inventore Charles K. Bliss). La particolare grafica e logica con cui sono composti i simboli, li rende infatti, idonei per essere utilizzati con soggetti molto diversi tra loro per livello intellettuale e quadro diagnostico.

I simboli Bliss sono segni grafici, basati sul significato e non sulla fonetica. Essi sono strutturati in un vocabolario standard e sono disponibili, sia sotto forma di francobolli adesivi, da collocarsi su supporti cartacei (tabelle di comunicazione), o in versione informatizzata. Il disabile può così comunicare sia attraverso l'indicazione dei simboli posti sulla propria tabella o la selezione degli stessi sul calcolatore: il calcolatore inoltre consente la trascrizione grafica e sonora dei messaggi.

L'utente in relazione all'età, livello intellettuale e possibilità motorie, può comporre messaggi di diversa complessità: da forme sintetiche simili alla parola o frase, a forme più articolate; ai simboli è sempre abbinata la traduzione in lingua per consentire a qualunque interlocutore di decodificare il messaggio.

Per il suo autore, il Bliss vuole essere uno strumento pratico, una scrittura ausiliaria a scopi comunicativi tra differenti nazioni ed inoltre, un mezzo per riconoscere significati vaghi e ambigui.

Bliss aveva considerato che solo una scrittura ideografica che avesse implicito il significato (semantografica) poteva, non solo, superare le diversità esistenti tra i linguaggi di ogni paese, ma anche quelle all'interno di una stessa lingua.

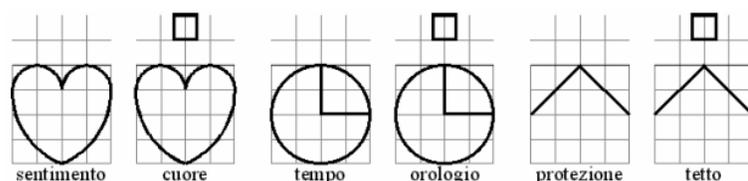
La differenza sostanziale tra una scrittura semantografica e la tradizionale scrittura con le 26 lettere, cioè tra una scrittura grafica e una fonetica è che la prima rappresenta i significati nel segno grafico stesso. Nella nostra mente il profilo della luna non evoca la combinazione di lettere l,u,n,a, ma immediatamente la cosa luminosa, d'argento, nel cielo, e questi significati li pensiamo nella nostra lingua madre.

Bliss ha quindi creato il suo linguaggio, ispirandosi alla scrittura ideografica cinese, che se inizialmente sembra difficile e complicata, una volta comprese l'organizzazione e la chiave di lettura risulta semplice e affascinante.

Il loro contenuto è sempre attuale perché deriva da una realtà esperita, che è comune all'uomo di ieri come all'uomo di oggi.

Per creare una scrittura di facile comprensione, Bliss ha tratto da quelle conoscenze di base comuni a tutti gli uomini gli elementi che costituiscono il vocabolario dei significati e li ha rappresentati con segni grafici. Tali segni, per la loro iconicità, potevano essere facilmente rapportati al significato rappresentato, o collegati a esso da una semplice deduzione logica.

Nel realizzare la struttura organizzativa del suo linguaggio, ha preso in considerazione ciò che accade nel mondo della natura, e ciò che fa parte di quello che viene definito senso comune, per trarne dei semplici principi di base che avessero una funzione sintattico grammaticale nel suo vocabolario di simboli.



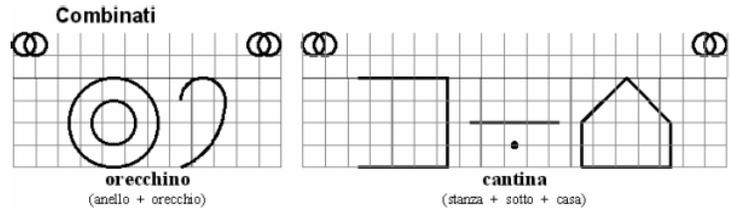


Figura 23: Esempi di simboli Bliss

Una delle critiche ricorrenti al sistema Bliss è la non immediata trasparenza dei significati. A questo problema è però possibile ovviare mediante un processo di caratterizzazione, ovvero la contestualizzazione grafica dell'aspetto semantico.

Il vero carattere distintivo dei codici alternativi è da ricercare nella possibilità comunicativa offerta: i simboli Bliss, sia per il loro aspetto combinatorio sia per la grafica geometrica, si prestano a rappresentare qualsiasi significato. Infatti è l'unico codice alternativo, insieme al Picsym, ad essere supportato da una ricca grammatica, con regole di composizione che permettono la scrittura di nuovi simboli, in modo tale da rendere possibile l'ampliamento del vocabolario, caratteristica fondamentale e irrinunciabile se si vuole che un sistema simbolico diventi un linguaggio di comunicazione.

Un secondo aspetto peculiare del Bliss consiste nella possibilità di rappresentare concetti astratti (ad es. amore, pace, lontano, vicino...). Negli altri codici, infatti, emerge la difficoltà di tradurre in forma iconica un significato astratto, in quanto si riscontra che tali linguaggi sono soprattutto dei dizionari di cui viene sviluppato in modo preponderante il lessico e non sono strutturati da regole formali organizzative.

Le precedenti esperienze in letteratura hanno dimostrato che l'uso del sistema Bliss, dal momento che è un sistema basato sul significato e non sulla fonetica, che implica nel suo uso delle regole di costruzione ed organizzazione (che possono essere semplificate o ridotte in base al grado di disabilità) ha spesso favorito non solo la comunicazione, ma anche l'apprendimento della lettura e della scrittura.

I simboli PCS invece sono semplici disegni, accompagnati o meno dalle parole corrispondenti al loro significato. I simboli PCS sono appropriati per persone che utilizzano un livello semplice di comunicazione. Ciò comporta un vocabolario limitato e una struttura morfosintattica non elaborata.

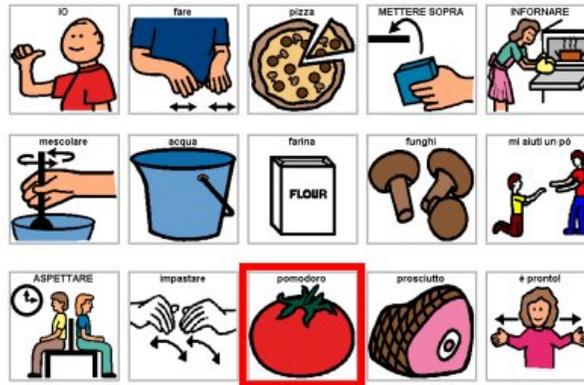


Figura 24: Esempi di procedure codificate con simboli PCS

I simboli PCS sono usati da insegnanti, terapisti, membri della famiglia per stimolare esperienze di linguaggio. È stata dimostrata l'efficacia dell'uso dei PCS per persone con limitata abilità di linguaggio verbale e/o limitate abilità di letto-scrittura.

Di seguito sono elencati alcuni dei benefici potenziali:

- I prerequisiti per la letto-scrittura possono essere stimolati dalla "lettura" dei simboli. La pratica di andare da sinistra verso destra e avere la parola scritta sul simbolo aumenta la possibilità di sviluppare le capacità di lettura.
- Si possono progettare tabelle per facilitare la comprensione e l'espressione del linguaggio. Gli ausili di comunicazione forniscono un'altra modalità per ricevere ed esprimere il linguaggio al di là della comunicazione verbale e scritta.
- La visualizzazione grafica può aiutare a compensare i deficit di memoria. Liste e istruzioni che usano simboli possono essere usati con chi non sia in grado di leggere.

L'uso dei simboli permette di imparare ed incrementare la sintassi (formazione di frasi e periodi). I simboli rafforzano attraverso la componente visiva il linguaggio che i lettori hanno.

3.3. Leggibilità

Nell'ambito della valutazione di un sito internet, la leggibilità è uno dei fattori determinanti per stabilire la cosiddetta "usabilità", ormai parola feticcio del mondo Internet. Per usability si intendono le condizioni di un efficace utilizzo di un sito.

Il tema della leggibilità è molto legato a internet: se il problema della leggibilità si pone per ogni testo che abbia uno scopo prevalentemente informativo, allora Internet entra a pieno titolo nella questione. Non si naviga infatti per avere delle informazioni? C'è di più. Se è

necessario essere chiari e semplici in qualsiasi documento pubblico, allora a maggior ragione bisogna esserlo in Internet. La scrittura, infatti, è vincolata al mezzo che la ospita. Farsi capire dal lettore è l'unico modo per sopravvivere nel Web. Qui il lettore è utente: usa il mezzo per i suoi scopi. Perciò è il lettore che decide cosa leggere, non lo scrittore cosa scrivere. Se su Internet "scrivere" deve diventare "scrivere per farsi capire", forse questa verità deve varcare i confini della Rete. Forse è l'occasione buona per indirizzare sulla via di una maggiore chiarezza la lingua italiana.

Pochi, comunque, sanno a cosa serve e in quale contesto si usi la parola leggibilità.

Intuitivamente possiamo capire che cosa significhi. È la condizione per cui un testo è leggibile, comprensibile. Leggibilità di un testo vuol dire facilità di lettura di quel testo. Intendiamoci. Il suo significato tecnico è più complesso, soprattutto nel campo Internet.

Innanzitutto, nel leggere un testo non va considerato solo l'aspetto linguistico, ma anche quello visivo. Esiste infatti una leggibilità **visiva**, che comprende tutti i problemi legati alla lettura su video. Nella maggior parte dei casi, un documento perfettamente leggibile su carta risulterà difficile da consultare online. Lo schermo del computer non è molto adatto alla lettura. La bassa risoluzione riduce la leggibilità. La fonte di luce proveniente dall'interno del monitor è affaticante per gli occhi. In genere, è stato verificato che in queste condizioni la lettura è più lenta del 25%-30% rispetto a quella su carta. Perciò si finisce per dar solo una rapida occhiata al testo. Poiché la modalità di lettura su schermo risulta differente da quella su carta è inutile inserire online così com'è un documento formattato per la stampa.

Accanto ad una leggibilità visiva c'è poi quella **linguistica**, che riguarda l'utilizzo della lingua in tutte le sue componenti, dalla scelta dei termini e della sintassi impiegata all'articolazione dei contenuti. Anche il problema della leggibilità linguistica riguarda essenzialmente quei testi caratterizzati da un obiettivo informativo. Di certo non leggiamo "I Promessi Sposi" solo per avere delle informazioni, ma per il gusto che ne ricaviamo. Non è lo stesso per il bugiardino di un farmaco o per un modulo della denuncia dei redditi, da cui ci aspettiamo istruzioni precise e comprensibili. Non ci dobbiamo meravigliare però se la leggibilità di una pagina del Manzoni e quella della maggior parte dei bugiardini si equivalgono.

Il fatto è che il nostro italiano presenta ancora un forte scarto tra scritto e parlato. Quanto il parlato è in continua evoluzione, tanto lo scritto rimane ancorato al modello conservatore di eredità scolastica e letteraria. Per non parlare poi del nostro gusto così italiano del complicato e dell'ampoloso. E la differenza con altre lingue moderne è spesso davvero evidente. Per

verificarlo, è sufficiente confrontare l'italiano con lo spirito pragmatico dell'inglese e con la prevalenza di costruzioni più lineari del francese.

Sicuramente si tratta di differenti opzioni di gusto. Così, quando scriviamo, noi italiani abbiamo la tendenza a dare ai nostri testi una veste ufficiale, formale. Li infarciamo di formule, giri di parole, stereotipi, tecnicismi e aulicismi. Negli Stati Uniti il problema della leggibilità dei testi scritti è in evidenza da più di un secolo. Non solo. È stato ritenuto a tal punto importante che 29 stati hanno firmato il "Plain Language Act". La "Legge sul parlar chiaro" impone alle aziende di scrivere con un 3 linguaggio chiaro e comprensibile qualsiasi tipo di messaggio. Inoltre, questa normativa stabilisce di usare come criterio di leggibilità l'indice Flesch. Tutti i documenti pubblici devono avere un indice di leggibilità non inferiore a 45. Questa legge, in realtà, è il risultato di anni di ricerca e di sensibilizzazione dell'opinione pubblica riguardo l'importanza della leggibilità. E' stato infatti lo studioso americano di origine austriaca Rudolph Flesch il primo a dichiarare che la leggibilità di un testo è un concetto quantitativo. E' stato, inoltre, il primo a proporre un metodo per misurarla, appunto l'Indice Flesch. In "The Art of Plain Talk" (1946) Flesch ha esposto il frutto di vent'anni di studi. In base a questi, un testo risulta difficile quando contiene molte subordinate (difficoltà sintattica) e molte parole astratte (difficoltà semantica). Esiste una correlazione tra parole astratte e numero di sillabe per campione, come tra proposizioni subordinate e numero medio di parole per periodo. Per questo Flesch ha assunto come indicatori di facilità il numero medio di parole per periodo in un campione di 100 parole e il numero di sillabe contenute in un campione di 100 parole. Ne ha ricavato la formula matematica che sta alla base dell'indice Flesch.

3.4. Gli indici di leggibilità: l'indice di Gulpease

Un indice di leggibilità è una formula matematica che attraverso un calcolo statistico è in grado di predire la reale difficoltà di un testo in base a una scala predefinita di valori.

Per definire la formula di un indice di leggibilità si può tener conto di diverse variabili linguistiche, cioè della misura di alcuni parametri del testo. Le variabili linguistiche più semplici sono, per esempio, lunghezza media delle parole e lunghezza media delle frasi.

Vi sono variabili linguistiche che sono indipendenti dal contenuto del testo, come appunto le due citate, oppure variabili linguistiche legate al lessico, alla struttura del periodo, ecc.

Sono state definite molte formule per la predizione della leggibilità, ma quelle che hanno avuto maggiore successo sono quelle che considerano variabili linguistiche di facile calcolo, come per esempio la lunghezza delle parole e delle frasi.

Per lo stesso motivo, i programmi che calcolano automaticamente il valore delle formule si limitano, nella maggior parte delle applicazioni, a formule semplici.

Nel 1987 un gruppo di linguisti dell'Università La Sapienza di Roma si è riunita attorno a Tullio de Mauro per costituire il GULP, gruppo universitario linguistico pedagogico. In questi laboratori linguistici è nato l'**Indice Gulpease**.

La formula per il calcolo dell'Indice Gulpease è stata determinata verificando con una serie di test la reale comprensibilità di un corpus di testi. La verifica è stata fatta su diversi tipi di lettore, e accanto alla determinazione della formula è stata definita una scala d'interpretazione dei valori restituiti dalla formula stessa.

L'**indice di Gulpease** è valutabile con la formula seguente (Lucidano – Piemontese 1988):

$$I_G = 89 - 10 \cdot \frac{N_L}{N_p} + 300 \cdot \frac{N_F}{N_p}$$

Dove N_L è il numero di lettere, N_p è il numero di parole, N_F è il numero di frasi. Si noti la proporzionalità diretta con il numero medio di frasi per parola, e la proporzionalità inversa con il numero medio di lettere per frase.

L'indice di Gulpease I_G risulterà quindi proporzionale alla facilità di lettura del testo preso in considerazione.

La **scala di leggibilità** secondo questo indice va da 100 (leggibilità massima) a 0 (leggibilità nulla) e mette in relazione i valori restituiti dalla formula con il grado di scolarizzazione del lettore.

- I lettori che hanno un'istruzione elementare leggono facilmente i testi che presentano un indice superiore a 80.
- I lettori che hanno un'istruzione media leggono facilmente i testi con indice superiore a 60.
- I lettori che hanno un'istruzione superiore leggono facilmente i testi con indice superiore a 40.

Un testo con indice Gulpease di circa 60 dunque è molto difficile per chi ha la licenza elementare, difficile per chi ha la licenza media, facile per chi ha un diploma superiore.

La formula Gulpease, oltre a essere la prima formula di leggibilità tarata direttamente sulla lingua italiana, ha anche il vantaggio di calcolare la lunghezza delle parole in lettere, e non in sillabe. Per ottenere una misura in tempo reale dell'indice di leggibilità, è stato creato un apposito strumento web che utilizza la formula Gulpease, che sarà messo a disposizione dei redattori di navigAble.

I testi inseriti nella sezione Informazione del sito pilota di NavigAble sono stati redatti in modo da avere un indice di leggibilità Gulpease tra 70 e 80 – sembrano ragionevolmente comprensibili – e, nella versione Basic, utilizzano soltanto termini appartenenti al dizionario di base della lingua Italiana del DeMauro.

Valore	Difficoltà di lettura
> 79	Molto facile
60 ÷ 79	Facile
50 ÷ 59	Difficile
35 ÷ 49	Molto difficile
< 35	Quasi incomprensibile

Tabella 2: Soglie di leggibilità corrispondenti all'indice di Gulpease

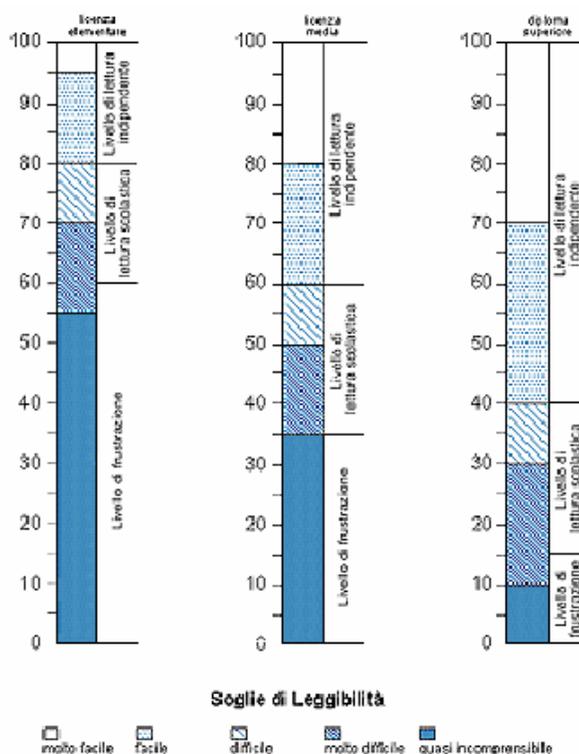


Figura 25: Indice di Gulpease: scale dei valori

L'entusiasmo per queste ricerche non si è dunque placato. In effetti continuano a nascere nuovi metodi per misurare sempre con maggiore precisione gli indici di leggibilità. Le ultime notizie arrivano dalla Francia, dove si stanno realizzando degli indici che tengano conto di criteri di analisi logica e del periodo. Ad esempio, i futuri indici non conteranno più solo il numero di parole per frase, ma anche le parole che dividono il verbo dal soggetto e il verbo dal complemento.

In questa direzione si muove una delle sfide del progetto di navigAbile. Nei paragrafi successivi verrà proposto un metodo di valutazione della leggibilità del testo basato sulla tipologia di costrutti sintattici che vengono utilizzati in un certo testo.

3.5. Plain Language: leggibilità e comprensibilità

Il Plain Language è definibile come il linguaggio che trasmette al lettore le informazioni in possesso dello scrittore nel modo più semplice ed efficace; idealmente il lettore dovrebbe riuscire a comprendere il testo così scritto alla prima lettura.

Il Plain Language altro non è che la lingua ordinaria, che si sforza di assomigliare a quella utilizzata nella conversazione quotidiana, senza tuttavia coincidere con essa, perché le differenze tra espressione scritta e orale non si possono annullare del tutto, né è desiderabile farlo.

Il Plain Language mira alla massima *leggibilità* e alla *comprensibilità* del testo. Anche se a prima vista può sembrare strano, un testo leggibile non è necessariamente comprensibile, e viceversa.

La **leggibilità** di un testo dipende dal modo in cui sono strutturate le sue frasi: in estrema sintesi, frasi brevi che contengono parole brevi sono più leggibili di frasi lunghe, così come frasi lineari sono più leggibili di frasi involute, ricche di incisi e subordinate.

La **comprensibilità** di un testo dipende dalla sua struttura, dalle parole che contiene, ma anche dalla condivisione di un universo culturale: gli specialisti di un settore specifico si comprendono perché condividono un contesto comune, che invece rimane estraneo ai profani. Di conseguenza, un testo molto specialistico potrebbe essere strutturalmente molto leggibile, ma comprensibile solo ad una cerchia ristretta di lettori: basti pensare ai testi di medicina quando vengono letti dai non addetti ai lavori. La comprensibilità è quindi collegata con quello che i linguisti chiamano leggibilità lessicale.

Nella scrittura di un testo destinato a un pubblico generico, è importante impiegare il più possibile una struttura di frase semplice e chiara per aumentare il grado di leggibilità sintattica. Questo però non esaurisce la questione: un ulteriore elemento che determina il grado di leggibilità di un testo è rappresentato dal grado di diffusione, o popolarità, delle parole che contiene. Un testo leggibile contiene parole ben conosciute dalla maggioranza dei lettori cui si rivolge.

Analisi statistico-linguistiche identificano il “vocabolario di base” (Il vocabolario di base dell’italiano, curato dal linguista Tullio De Mauro, è composto di 7050 parole, comprensibili al 60% circa della popolazione), cioè il gruppo di parole che costituisce il repertorio basilare di una data lingua. Nella scrittura di un testo destinato a un pubblico generico, è importante impiegare il più possibile parole che appartengano al vocabolario di base, per aumentare il grado di leggibilità lessicale. In particolare, per quanto riguarda un servizio di home banking, è importante che esso contenga testi che rassicurino il cliente e che consentano una tranquilla fruibilità dei contenuti: trattandosi di un servizio bancario l’utente di fatto pone già una tensione medio-alta nell’utilizzo del sistema.

Dato che la leggibilità di un testo dipende in primo luogo dalla struttura delle sue frasi, dalla lunghezza di queste ultime e delle parole che le compongono, si può dire che la leggibilità dipende da fattori quantitativi ed è quindi misurabile. Gli indici di leggibilità rappresentano gli strumenti di misura della leggibilità di un testo e pertanto definiscono una scala e dei valori, grazie ai quali è possibile confrontare la leggibilità di testi differenti.

Gli indici di leggibilità più utilizzati per la lingua italiana sono l’indice Flesch-Vacca e l’indice Gulpease. Come tutti gli strumenti automatici, questi indici valutano il profilo puramente quantitativo e non qualitativo: possono pertanto fornire indicazioni utili, ma vanno affiancati da un test con un campione rappresentativo (con caratteristiche socio-culturali simili) di destinatari reali del servizio per verificarne il reale grado di comprensione.

3.6. Vocabolario fondamentale di De Mauro

Il **Vocabolario di Base della lingua italiana** (VdB) di Tullio De Mauro, pubblicato nel 1997, è un elenco di lemmi elaborato prevalentemente secondo criteri statistici. Esso rappresenta la porzione della lingua italiana usata e compresa dalla maggior parte di coloro che parlano italiano.

La scelta dei lemmi è stata fatta in base ai primi 5.000 lemmi del “Lessico Italiano di Frequenza” (LIF) (ridotti a circa 4.750 dopo averne verificato la comprensibilità), integrati con un insieme di lemmi determinati per altre vie. In particolare, i lemmi del VdB sono classificati in tre livelli:

- **Vocabolario fondamentale:** i primi 1.991 lemmi del LIF. Sono i lemmi più frequenti in assoluto della nostra lingua;
- **Vocabolario di alto uso:** i successivi 2.750 lemmi dell'insieme preso dal LIF. Sono lemmi ancora molto frequenti, anche se molto meno di quelli del vocabolario fondamentale;
- **Vocabolario di alta disponibilità:** 2.337 lemmi determinati in vario modo, soprattutto con dizionari dell'italiano comune. L'integrazione è stata necessaria perché il LIF è il risultato dello spoglio di testi scritti, ed è ormai dimostrato che vi è in tutte le lingue un insieme di lemmi che, pur essendo quasi del tutto assenti nella lingua scritta, sono a tutti noti. Per esempio, lemmi come “forbice”, “abbronzare”, ecc.

Il VdB è stata la prima opera di questo genere realizzata in Italia, e a tutt'oggi è uno strumento d'importanza fondamentale per controllare e migliorare la leggibilità di un testo secondo criteri scientifici. Infatti le parole che non sono nel VdB sono meno comprensibili per le persone meno scolarizzate o quelle poco abituate a leggere, e quindi se si vuole essere sicuri che il testo sia comprensibile a tutti bisognerebbe usare solo tali parole.

Questo non vuol dire che un testo ad alta leggibilità sia chiuso a nuove parole, ma vuol dire che quando si usa un lemma non di base, soprattutto quando è un lemma importante per capire il testo, bisogna spiegarne il significato, usando nella spiegazione solo lemmi presenti nel VdB. Un testo altamente comprensibile non ha solo questa caratteristica, ma l'uso del VdB è condizione comunque necessaria.

Un esempio di scrittura ad alta leggibilità è il mensile “**AltaFrequenza**” (www.altafrequenza.it), edito dal dipartimento di Scienze del Linguaggio dell'Università di Roma «La Sapienza». I redattori di AltaFrequenza scrivono articoli usando il VdB e spiegando le parole che non vi appartengono. Inoltre, la redazione applica anche altri criteri di leggibilità, attinenti alla grafica (caratteri grandi, illustrazioni che non spezzano il testo, ecc.) e al modo di disporre i contenuti (riquadri di spiegazione, testo a nuova riga per ogni frase, ecc.).

3.6.1. L'indicatore di comprensibilità di navigAbile

L'indicatore di comprensibilità utilizzato nel contesto del progetto navigAbile tiene conto della percentuale del numero di parole appartenenti al vocabolario di base; in particolare vengono assegnati opportuni pesi alle parole che appartengono ai diversi livelli del vocabolario di base e vengono rapportati al numero di parole totale; si tratta quindi di un indicatore basato su una conoscenza lessicale.

L'indicatore di comprensibilità I_{COMP} introdotto assume la seguente espressione:

$$I_{COMP} = \frac{\alpha \cdot N_f + \beta \cdot N_u + \gamma \cdot N_d}{N} \cdot 100$$

Dove N è il numero di parole totali del testo preso in considerazione, N_f è il numero di parole appartenenti al vocabolario fondamentale, N_u è il numero di parole appartenenti al vocabolario ad alto uso, mentre N_d è il numero di parole appartenenti al vocabolario ad alta disponibilità. Tale indicatori assume valori nell'intervallo $0 \div 100$; il suo valore risulta proporzionale al livello di comprensibilità del testo a cui si riferisce.

Infine α, β, γ sono dei pesi che assumono valori decrescenti per tener conto della diversa comprensibilità dei tre livelli del vocabolario di base. Nel contesto del progetto navigAbile sono stati impiegati i seguenti valori per i pesi: $\alpha = 1, \beta = 0,75, \gamma = 0,5$. Per dare un'interpretazione intuitiva dell'indicatore di comprensibilità basta ricordare che se si considera un testo contenente esclusivamente parole appartenenti al vocabolario fondamentale e parole che non appartengono al VdB, l'indicatore di comprensibilità rappresenta proprio la percentuale di parole che appartengono al vocabolario fondamentale, con riferimento al numero totale di parole. Si può quindi pensare di stabilire due soglie, che individuano tre intervalli nell'intervallo $0 \div 100$, a cui possiamo associare i tre livelli di semplificazione del testo. Tuttavia un meccanismo a soglie fisse è troppo rigido. Il valore dell'indicatore di comprensibilità varia molto a seconda del contesto della notizia che si sta considerando. Quindi l'informazione derivante dall'indicatore di comprensibilità può essere utilmente sfruttata in sede di valutazione dei 3 livelli di semplificazione del testo, verificando che l'indicatore abbia un andamento crescente passando dal livello *Normal*, al livello *Basic*.

Ricordo infine che l'informazione sulla comprensibilità non è sufficiente per la valutazione di un testo, e deve essere integrata con un indicatore di leggibilità.

4. Il dizionario Multimodale di navigAbile

4.1. Struttura e schema concettuale

In generale, le entità presenti nella base di dati di navigAbile possono essere raggruppate in 3 macro categorie:

1. entità relative al vocabolario multimodale
2. entità relative all'inserimento di frasi composte
3. entità relative ai profili di utenza

Di seguito viene riportato lo schema logico della base di dati, relativo alle entità del vocabolario multimodale.

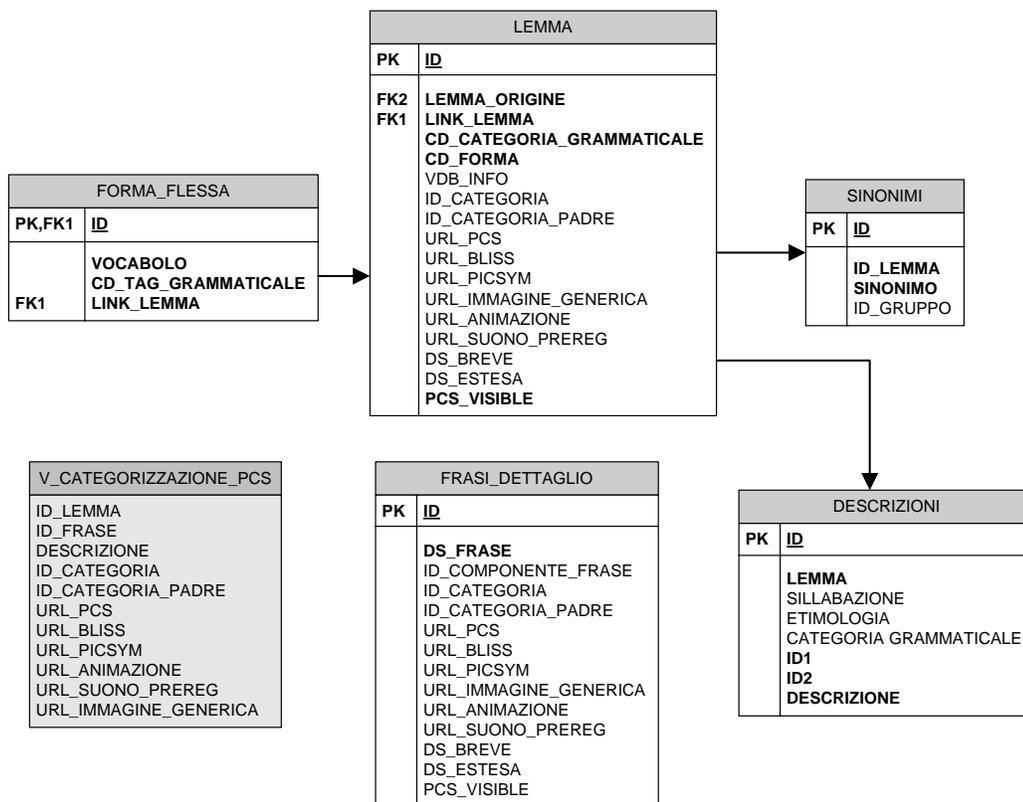


Figura 26: Schema delle tabelle della base di dati relativa al dizionario di navigAbile

Con riferimento al vocabolario multimodale possiamo individuare le seguenti entità:

FORMA_FLESSA: è la tabella che contiene tutte le forme flesse dei lemmi della lingua italiana. Ad esempio per il lemma “andare” sono presenti le forme flesse: “vado”, “vai”, etc. Per ogni forma flessa è presente una tag grammaticale che permette di ottenere la forma

flessa a partire dal lemma origine: ad esempio per la forma flessa “vado”, la tag grammaticale è “Indicativo presente, prima persona singolare”. Ogni forma flessa ha poi associato un lemma origine, attraverso una relazione con l’entità “LEMMA”: ogni forma flessa ha quindi un attributo (“LINK_LEMMA”) che specifica l’omonimo attributo della tabella “LEMMA” a cui è associata la forma flessa. L’attributo “VOCABOLO” contiene la forma flessa vera e propria. La chiave primaria è il contatore “ID”.

Attributo	Tipo	PK-Req'd	Descrizione
ID	LONG	PK	Contatore
VOCABOLO	VARCHAR(30)	Req'd	Contiene la forma flessa vera e propria. Es.: “vado”
CD_TAG_GRAMMATICALE	SHORT	Req'd	TAG codificato che include informazioni per ottenere la forma flessa a partire dal lemma origine. Ad esempio per “abbagliano” la tag vale “Indicativo, Presente, 3 pers. Singolare”
LINK_LEMMA	LONG	Req'd	LINK_LEMMA è un puntatore all’omonimo attributo nella tabella LEMMA

LEMMA: questa tabella contiene tutti i lemmi della lingua italiana. La chiave primaria non è il lemma stesso (infatti la stessa parola può appartenere a categorie grammaticali diverse) ma un ID progressivo. I lemmi uguali hanno lo stesso attributo “LINK_LEMMA” che viene usato per la relazione con la tabella “FORMA_FLESSA”. Ad esempio la forma flessa “abbagliante” punta, attraverso “LINK_LEMMA” ai due lemmi origine omonimi, corrispondenti al sostantivo e al verbo. L’attributo “CD_CATEGORIA_GRAMMATICALE” rappresenta la codifica della categoria grammaticale del lemma, come ad esempio “Sostantivo M”. L’attributo “VDB_INFO” codifica attraverso un carattere l’appartenenza o meno al Vocabolario di Base di De Mauro: F indica l’appartenenza al vocabolario fondamentale, U indica l’appartenenza al vocabolario ad alto uso, D indica l’appartenenza al vocabolario ad alta disponibilità, mentre qualsiasi altro carattere indica la non appartenenza al VdB. L’attributo “CD_FORMA” codifica la forma della categoria grammaticale, ad esempio “Maschile singolare”. “ID_CATEGORIA” individua la categoria rappresentata dal lemma, nel caso in cui esso sia contenuto nell’albero di categorizzazione in un nodo intermedio, mentre “ID_CATEGORIA_PADRE” individua la categoria di appartenenza. Sono presenti poi 5 attributi (“ID_BLISS”, “ID_PCS”, “ID_ANIMAZIONE”, “ID_SUONO_PREREG”, “ID_IMMAGINE_GENERICA”) che rappresentano puntatori alle relative tabelle, contenti gli elementi multimediali associati al lemma e 2 attributi

("DS_BREVE" e " DS_ESTESA") che contengono due definizioni del termine, una sintetica e una più dettagliata. La chiave primaria è il contatore "ID".

Attributo	Tipo	PK-Req'd	Descrizione
ID	LONG	PK	Contatore
LEMMA_ORIGINE	VARCHAR(30)	Req'd	Contiene la forma flessa vera e propria
LINK_LEMMA	LONG	Req'd	LINK puntato da una forma flessa ambigua (più lemma associati ad una forma flessa)
CD_CATEGORIA_GRAMMATICALE	SHORT	Req'd	Categoria grammaticale da codificare (es. aggettivo, sostantivo)
CD_FORMA	SHORT	Req'd	Forma grammaticale da codificare (es. maschile singolare per un sostantivo o indicativo presente per un verbo)
VDB_INFO	CHAR(1)		Livelli Vocabolario di Base (F: Fondamentale; U: Alto Uso; D: Alta Disponibilità). Se non specificato non appartiene al vocabolario di base.
ID_CATEGORIA	LONG		ID della categoria rappresentata dall'oggetto
ID_CATEGORIA_PADRE	LONG		ID della categoria a cui appartiene l'oggetto
URL_PCS	VARCHAR(200)		URL che punta all'immagine PCS associata
URL_BLISS	VARCHAR(200)		URL che punta all'immagine Bliss associata
URL_PICSYM	VARCHAR(200)		URL che punta all'immagine PICSYM associata
URL_ANIMAZIONE	VARCHAR(200)		URL che punta all'animazione associata
URL_SUONO_PREREG	VARCHAR(200)		URL che punta al suono preregistrato associato
URL_IMMAGINE_GENERICA	VARCHAR(200)		URL che punta all'immagine generica associata
DS_BREVE	VARCHAR(50)		Descrizione sintetica del significato del lemma
DS_ESTESA	VARCHAR(200)		Descrizione dettagliata del significato del lemma
PCS_VISIBLE	BIT	Req'd	Flag che indica se l'immagine PCS appartiene al set di 150 immagini di cui si ha la licenza Mayer (in questo caso PCS_VISIBLE=1) o meno. I profili di prova possono vedere solo il set di 150 immagini.

SINONIMI: questa tabella raccoglie i sinonimi associati ad ogni lemma. L'attributo "SINONIMO" contiene un lemma che è sinonimo del lemma di origine a cui è associato; tale associazione avviene attraverso l'attributo "ID_LEMMA", che punta alla chiave primaria della tabella "LEMMA". Per ogni lemma origine possono esistere più sinonimi, che possono essere raggruppati in gruppi di significato diverso: l'attributo "ID_GRUPPO" serve per raggruppare i sinonimi con lo stesso significato. La chiave primaria è il contatore "ID".

Attributo	Tipo	PK-Req'd	Descrizione
ID	LONG	PK	Contatore
ID_LEMMA	LONG	Req'd	ID puntatore alla PK della tabella LEMMA
SINONIMO	VARCHAR(30)	Req'd	Contiene il sinonimo vero e proprio
ID_GRUPPO	SHORT		ID_GRUPPO è un identificativo che raggruppa per insiemi di concetto: i sinonimi con lo stesso ID_LEMMA sono raggruppati in un certo numero di gruppi, e ID_GRUPPO ne è l'identificativo

Sono riportati di seguito degli esempi esplicativi dei dati contenuti nelle tabelle sopra descritte:

FORMA_FLESSA

- 1;abbagliante;Maschile singolare;(3)abbagliante;
- 2;abbaglianti;Maschile plurale;(3)abbagliante;
- 3;abbagliante;Femminile singolare;(3)abbagliante;
- 4;abbaglianti;Femminile plurale;(3)abbagliante;
- 5;abbagliante;Singolare;(4)abbagliante;
- 6;abbaglianti;Plurale;(4)abbagliante;
- 7;abbaglio;Indicativo, Presente, 1 pers. singolare;(5)abbagliare;
- 8;abbagli;Indicativo, Presente, 2 pers. singolare;(5)abbagliare;
- 9;abbaglia;Indicativo, Presente, 3 pers. singolare;(5)abbagliare;

Nota: nel campo CD_TAG_GRAMMATICALE è presente la codifica della tag grammaticale e non la descrizione della stessa. Qui si è riportato direttamente la descrizione per maggiore chiarezza espositiva.

LEMMA

- 1;3;abbagliante;Aggettivo;D;Maschile singolare;
- 2;3;abbagliante;Aggettivo;D;Femminile singolare;
- 3;4;abbagliante;Sostantivo M;D;Singolare;
- 4;5;abbagliare;Verbo 1^a Tr;U;Infinito, Presente;

Nota: nei campi CD_CATEGORIA_GRAMMATICALE e CD_FORMA è presente la codifica delle categorie grammaticali e le forme grammaticali e non la descrizione delle stesse.

SINONIMI

1;casa;abitazione;1

2;casa;dimora;1

DESCRIZIONI: la tabella contiene le descrizioni formali dei lemmi; tali descrizioni sono state ricavate sulla base del dizionario Garzanti online. I campi ID1 e ID2 sono degli identificatori incrementali servono per specificare il livello di aggregazione dei diversi significati dello stesso vocabolo. Servono nel caso in cui un vocabolo abbia più significati: ID1 serve per raggruppare vocaboli con significati simili, mentre ID2 serve per discriminare i diversi significati dei vocaboli con lo stesso valore di ID1. Al momento questa entità non è utilizzata.

Attributo	Tipo	PK-Req'd	Descrizione
ID	LONG	PK	Identificatore univoco della descrizione
LEMMA	VARCHAR(255)	Req'd	E' il lemma a cui si riferisce la descrizione
SILLABAZIONE	VARCHAR(255)	Req'd	Riporta la suddivisione in sillabe del vocabolo, ad esempio [ab-ban-do-nà-re]
ETIMOLOGIA	VARCHAR(255)	Req'd	Spiegazione dell'origine del lemma
CATEGORIA GRAMMATICALE	VARCHAR(255)	Req'd	Il campo specifica la categoria grammaticale del vocabolo; tale campo non è sempre presente
ID1	SHORT	Req'd	Identificativo che discrimina lemmi uguali aventi significati diversi
ID2	SHORT	Req'd	Identificativo che discrimina lemmi uguali aventi significati diversi
DESCRIZIONE	VARCHAR(370)	Req'd	Descrizione del lemma; si tratta della spiegazione che si legge sul vocabolario

FRASI_DETAGLIO: questa tabella contiene le frasi che corrispondono a concetti che possono essere espressi tramite un unico oggetto di ausilio (immagine PCS, immagine bliss, animazione o altro): l'attributo "DS_FRASE" contiene la descrizione della frase (cioè la frase vera e propria), mentre i campi rimanenti svolgono una funzione analoga agli omonimi contenuti nella tabella LEMMA: i campi relativi alle categorie ("ID_CATEGORIA" e "ID_CATEGORIA_PADRE") consentono di collocare l'oggetto all'interno dell'albero di navigazione, i campi relativi alle URL consentono di recuperare le forme alternative in cui può

essere espressa la frase utilizzando altri linguaggi, i campi descrizione – infine – provvedono a chiarificare il significato dell'espressione.

Attributo	Tipo	PK-Req'd	Descrizione
ID	LONG	PK	Contatore
ID_COMPONENTE_FRASE	LONG		???
DS_FRASE	VARCHAR(100)	Req'd	Contiene la descrizione della frase
ID_CATEGORIA	LONG		ID della categoria rappresentata dall'oggetto
ID_CATEGORIA_PADRE	LONG		ID della categoria a cui appartiene l'oggetto
URL_PCS	VARCHAR(200)		URL che punta all'immagine PCS associata
URL_BLISS	VARCHAR(200)		URL che punta all'immagine Bliss associata
URL_PICSYM	VARCHAR(200)		URL che punta all'immagine PICSYM associata
URL_ANIMAZIONE	VARCHAR(200)		URL che punta all'animazione associata
URL_SUONO_PREREG	VARCHAR(200)		URL che punta al suono preregistrato associato
URL_IMMAGINE_GENERICA	VARCHAR(200)		URL che punta all'immagine generica associata
DS_BREVE	VARCHAR(50)		Descrizione sintetica del significato della frase
DS_ESTESA	VARCHAR(200)		Descrizione dettagliata del significato della frase
PCS_VISIBLE	BIT	Req'd	Flag che indica se l'immagine PCS appartiene al set di 150 immagini di cui si ha la licenza Mayer (in questo caso PCS_VISIBLE=1) o meno. I profili di prova possono vedere solo il set di 150 immagini.

V_CATEGORIZZAZIONE_PCS: questa vista integra tutti i record presenti nelle tabelle LEMMA e FRASI_DETAGLIO, catalogati nell'albero di categorizzazione. In questo modo è possibile navigare per categorie il dizionario interrogando solamente la vista, pur mantenendo ben distinti in memoria i lemmi dalle frasi. I campi "ID_LEMMA" e "ID_FRASE" corrispondono all'ID nella tabella corrispondente all'oggetto, la "DESCRIZIONE" replica "LEMMA_ORIGINE" o "DS_FRASE", mentre i campi rimanenti hanno lo stesso nome e la stessa funzione in entrambe le tabelle.

Attributo	Descrizione
ID_LEMMA	ID importato dalla tabella LEMMA

ID_FRASE	ID importato dalla tabella FRASE
DESCRIZIONE	Contiene la descrizione della frase o del lemma
ID_CATEGORIA	ID della categoria rappresentata dall'oggetto
ID_CATEGORIA_PADRE	ID della categoria a cui appartiene l'oggetto
URL_PCS	URL che punta all'immagine PCS associata
URL_BLISS	URL che punta all'immagine Bliss associata
URL_PICSYM	URL che punta all'immagine PICSYM associata
URL_ANIMAZIONE	URL che punta all'animazione associata
URL_SUONO_PREREG	URL che punta al suono preregistrato associato
URL_IMMAGINE_GENERICA	URL che punta all'immagine generica associata

Il DBMS scelto per l'implementazione del database è Microsoft SQL Server 2000 (SP3).

4.2. Scenari di utilizzo tipici del dizionario di navigAbile

Le funzionalità principali di navigAbile che richiedono l'impiego delle dizionario multimodale sono:

- Navigazione visuale del dizionario di navigAbile. Esistono due tipologie di navigazione del dizionario: per categorie (Dizionario Categorie) e per ordine alfabetico (Dizionario Alfabetico). Nel caso della navigazione per categorie viene eseguita una query sulla tabella LEMMA, che seleziona tutti i lemmi che appartengono alla categoria che viene visualizzata in quel momento. Nel caso invece di navigazione per ordine alfabetico, vengono selezionati tutti i lemmi che iniziano con una certa lettera.
- Visualizzazione di una voce del dizionario. Questa funzione viene invocata all'interno delle pagine di contenuto: i vocaboli meno ricorrenti (quelli non appartenenti al VdB) vengono sottolineati con un collegamento ipertestuale che rimanda a una pagina contenente la descrizione del vocabolo, che è un attributo del lemma.
- Supporto all'analisi sintattica. Ogni lemma ha nei suoi attributi delle informazioni che specificano la sua categoria grammaticale. Questo tipo di informazioni vengono utilizzate dall'analizzatore sintattico del testo.
- Supporto al calcolo degli indici di comprensibilità e leggibilità. Ogni lemma ha un attributo che specifica il livello di appartenenza al vocabolario di base, ed è quindi possibile calcolare l'indice di comprensibilità. Durante la valutazione di questo indice, viene eseguita una query che esegue una join tra le tabelle FORMA_FLESSA e LEMMA,

in modo da selezionare il lemma origine corrispondente ad ogni forma flessa, e infine si controlla a quale livello del dizionario appartiene il lemma origine. Eseguendo questa operazione per tutte le parole che sono presenti nel testo, permette di valutare l'indice di comprensibilità lessicale del testo considerato.

- Servizio di suggerimento dei sinonimi (thesaurus): quando un lemma di un certo non appartiene al vocabolario fondamentale (il set più ristretto di vocaboli), vengono ricercati suoi eventuali sinonimi che appartengono ad un livello più ristretto del vocabolario di base di De Mauro; i lemmi che soddisfano queste condizioni vengono proposti al redattore, il quale può scegliere il sinonimo che meglio si adatta al contesto della notizia.

5. Analisi di testi

5.1. Introduzione alla comprensione del linguaggio

In generale, lo scopo di un programma per la comprensione del linguaggio naturale consiste nella traduzione delle frasi in una qualche rappresentazione interna sulla quale sia possibile compiere, in modo automatico, inferenze e ragionamenti. Questo compito è tutt'altro che semplice: i linguaggi naturali sono spesso ambigui, consentono di esprimere una grande varietà di sfumature di significato e nel servirsene ogni essere umano fa inconsapevolmente uso di un gran numero di conoscenze.

La ricchezza del vocabolario, la flessibilità della sintassi e le diverse forme di ambiguità sono alcune delle proprietà della lingua che contribuiscono a rendere l'analisi automatica del linguaggio un'operazione piuttosto complessa.

Bisogna sottolineare che la trasformazione del testo iniziale in una rappresentazione interna richiede l'utilizzo di molteplici fonti di conoscenza: da quella fonologica (rilevante esclusivamente per il riconoscimento del parlato) a quella pragmatica, passando attraverso successive fasi di analisi morfologica, lessicale, sintattica e semantica.

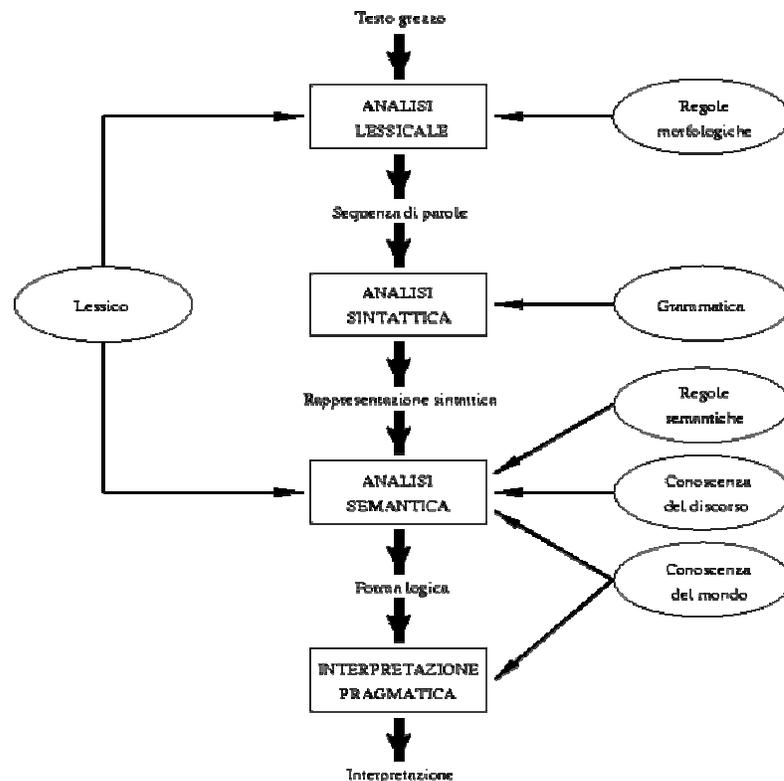


Figura 27: Il processo di analisi di un testo nell'elaborazione del linguaggio naturale

Nel processo di analisi si possono individuare diverse fasi logicamente distinte (vedi figura):

- **Analisi lessicale:** in primo luogo occorre suddividere la sequenza di caratteri che costituisce il testo grezzo in corrispondenza dei punti di separazione tra le parole, dei segni di punteggiatura e dei vari simboli grafici che possono comparire. Le informazioni sintattiche e semantiche delle parole sono contenute nelle corrispondenti voci del lessico, una delle fonti di conoscenza a disposizione del sistema. Se nel lessico vi è una voce per ciascuna forma flessa (una per "mangio", una per "mangi", una per "mangia" e così di seguito) sarà possibile accedervi direttamente a questo punto dell'analisi. Se invece nel lessico vi sono voci solo per le forme di citazione (l'infinito dei verbi, il singolare maschile per gli aggettivi ecc.) o per le radici (per esempio "mangi"), allora è necessaria una analisi morfologica preliminare. In linguistica, per indicare le unità minime di costruzione delle parole di una lingua si fa ricorso al concetto di "morfema": i morfemi non possono essere ulteriormente scissi in parti riconoscibili o dotate di significato. Semplificando, si può dire che l'analisi morfologica consiste nella suddivisione di ciascuna parola in una radice, ed eventualmente in uno o più affissi, ovvero segmenti collocati all'inizio (prefissi) o alla fine (suffissi) della radice, attraverso cui si specifica ulteriormente il significato di una parola (così, ad esempio, l'affisso "-are" connesso alla radice del verbo "mangi-" individua la forma di modo infinito). Nel caso di alcune lingue questa analisi può essere ulteriormente complicata dalla presenza di infissi, morfemi che si aggiungono all'interno di un altro morfema. L'analisi morfologica può essere anche molto complessa, specie per le lingue fortemente irregolari, ma consente di evitare la notevole ridondanza di un lessico con una voce per ciascuna delle forme flesse.
- Una volta associato un corredo di informazioni a ciascuna delle parole si può passare a prendere in considerazione le frasi. La conoscenza sintattica del sistema, quella che consente di stabilire se un determinata stringa in *input* costituisce una frase corretta nella lingua considerata, è mantenuta in una "grammatica", una rappresentazione formale delle strutture accettabili di un linguaggio. L'analizzatore sintattico, o *parser*, si serve della grammatica per generare, nel caso in cui l'*input* sia corretto, una rappresentazione della struttura delle frasi. In realtà, in generale, a ciascuna frase potrà essere associata più di una struttura a causa dell'ambiguità intrinseca in ogni lingua naturale.

-
- Ottenuta la struttura sintattica della frase, servendosi delle informazioni associate al significato di ciascuna delle parole che vi compaiono si può passare all'analisi semantica. In molti casi, il risultato di questa analisi è una forma logica, un'espressione generalmente traducibile in una formula del calcolo dei predicati, che costituisce una rappresentazione delle condizioni di verità della frase isolata dal contesto. Nel caso in cui sia presente un riferimento ad un elemento esterno alla frase stessa (ad esempio un pronome), la forma logica sarà quindi una rappresentazione incompleta in quanto al ruolo sintattico coperto dal pronome sarà possibile associare un concetto solo prendendo in considerazione anche le frasi precedenti. Ad esempio, la frase "Dillo a lui" ha bisogno di un contesto per essere interpretata completamente: la frase isolata non consente di stabilire chi si intende con "lui".
 - Il ruolo della "pragmatica" consiste nella determinazione definitiva del significato del testo nel suo complesso. Si tratta di un compito che richiede non solo una padronanza dei fenomeni linguistici e delle regole che governano la struttura del discorso ma anche una conoscenza dettagliata del dominio nell'ambito del quale il sistema è chiamato a comunicare.

Nella maggior parte dei sistemi moderni, considerazioni relative all'efficienza spingono ad intrecciare le diverse fasi descritte; queste devono quindi considerarsi distinte solo da un punto di vista logico.

5.2. L'analisi lessicale

Tra i passi che costituiscono il trattamento di una frase, l'analisi lessicale è quello apparentemente più semplice e, forse per questa ragione, il più trascurato dalla letteratura sulla linguistica computazionale.

In primo luogo occorre separare tra loro le parole, i segni di punteggiatura e i simboli che compaiono nel testo. I simboli grafici e di punteggiatura vengono trattati, nelle fasi successive, come categorie lessicali particolari.

Nel caso in cui ci si serva di un analizzatore morfologico, si procede tentando di scomporre ciascuna parola nei morfemi che la compongono. Ad esempio, se si è incontrata la parola "rimangiammo" si deve pervenire alla scomposizione in un prefisso "ri-", che indica che si

tratta di un'azione ripetuta, una radice “-mangi-” che individua il verbo, e un suffisso “-ammo”, che stabilisce che si deve considerare la prima persona plurale del passato remoto.

Se si accetta l'ipotesi semplificativa secondo cui i morfemi possono essere composti unicamente per concatenazione è possibile eseguire la decomposizione con dei riconoscitori a stati finiti. L'assunzione che l'unica modalità di combinazione dei morfemi sia la concatenazione è ragionevole per molte lingue. Per quanto riguarda l'italiano, si rende necessario un trattamento specifico per i verbi irregolari, che si riduce in genere all'utilizzo di più radici, ciascuna delle quali può supportare solo un sottoinsieme delle desinenze.

In molti sistemi, tuttavia, si è preferito aggirare la complessità aggiuntiva introdotta da un analizzatore morfologico ricorrendo a lessici in cui le voci sono associate direttamente alle forme flesse. È evidente che in questo caso le voci sono molto più numerose, e si introduce una ridondanza notevole dovendo ripetere molte informazioni.

Nelle applicazioni reali non è in genere lecito assumere che il lessico contenga tutte le parole che è possibile incontrare. In particolare, esistono classi di parole, come ad esempio i nomi propri, che rendono necessario un trattamento specifico. Le analisi indirizzate al trattamento di questo problema si sono per lo più concentrate sullo sviluppo di regole euristiche per l'attribuzione delle etichette lessicali basate sulle caratteristiche morfologiche della parola sconosciuta, sull'impiego delle lettere maiuscole e sull'analisi delle classi lessicali delle parole vicine.

Per quanto riguarda l'informazione semantica associata a ciascuna parola, la necessità di avere un lessico con la migliore copertura possibile e di mantenere allo stesso tempo una descrizione del dominio gestibile ha condotto in alcuni casi all'impiego di dizionari a due livelli: per i termini considerati più significativi, o che compaiono frequentemente nel dominio di applicazione che il sistema è destinato a trattare, si riporta una entrata lessicale completa di tutte le informazioni semantiche, mentre per le altre parole si ricorre alla creazione di un dizionario più generico in cui sono riportate solo informazioni sintattiche superficiali.

Indipendentemente dalla tecnica impiegata per la ricerca lessicale, occorre inoltre progettare dei metodi per il riconoscimento delle date e delle quantità espresse tramite numeri: in genere sono sufficienti dei semplici riconoscitori a stati finiti.

Al termine dell'analisi lessicale a ciascuna parola potrà essere associata non una singola entrata lessicale, ma un insieme di tali entrate, sia perché in ogni lingua le parole hanno spesso più di un significato, sia perché può accadere che una stessa forma flessa possa essere

interpretata in più modi diversi (ad esempio, “pesca” non solo presenta l'ambiguità tra l'interpretazione come sostantivo e quella come voce del verbo “pescare”, ma anche quella tra diverse voci del verbo quali l'imperativo e l'indicativo).

Nell'ambito della linguistica computazionale l'ambiguità è una delle sorgenti principali di complessità computazionale, ed è perciò particolarmente importante che essa sia per quanto possibile ridotta. Una strada che è possibile seguire consiste nel sottoporre il testo ad una operazione di “etichettatura delle parti del discorso” (*part of speech tagging*) attraverso cui si individuano delle preferenze per particolari interpretazioni delle parole ambigue servendosi di conoscenze linguistiche e di informazioni statistiche (ad esempio, il calcolo della probabilità di occorrenza di una parola in un determinato contesto), ma senza eseguire un'analisi sintattica vera e propria. In particolare, l'etichettatura stabilisce una preferenza per una certa categoria lessicale tra quelle possibili per una parola: nel caso di “pesca” non aiuta a scegliere tra l'imperativo e la terza persona singolare dell'indicativo presente del verbo “pescare”, ma può stabilire che in un particolare contesto è più probabile che si tratti di una voce del verbo “pescare” piuttosto che di un sostantivo. La difficoltà principale nella realizzazione di questo tipo di etichettatura risiede nella estrazione delle probabilità, sia quelle di tipo “lessicale”, associate alle singole parole, sia quelle di tipo “contestuale”, cioè associate alle relazioni con altre parole e parti del discorso. Il metodo più diffuso per superare queste difficoltà richiede la disponibilità di un *corpus* già etichettato di testi analoghi a quelli con cui il sistema si dovrà confrontare (i *corpora*, di cui si parlerà diffusamente nel capitolo 4, sono raccolte di testi rappresentativi del linguaggio corrente): le probabilità delle assegnazioni vengono approssimate tramite le frequenze corrispondenti.

5.3. L'analisi sintattica

Lo scopo dell'analisi sintattica consiste nella verifica della correttezza delle frasi e nella determinazione della loro struttura. Questa fase del processo di comprensione non è ancora direttamente coinvolta nell'analisi del significato. A questo livello di analisi occorre tener presente la distinzione tra “buona formazione dal punto di vista sintattico” e “significato”: l'analisi sintattica non deve cioè indagare il rapporto delle varie espressioni linguistiche con i loro significati ma individuare la struttura sintattica delle frasi in input e verificarne la correttezza. Semplificando, si può indicare la sintassi come quella parte della linguistica che si

occupa delle combinazioni possibili delle parole e delle leggi che governano tali combinazioni. Per mostrare che la buona formazione (in senso sintattico) non ha nulla a che fare col significato, ma è organizzata secondo leggi proprie, si può ricorrere alla ormai classica coppia di esempi dovuta a Chomsky:

(1) *Colorless green ideas sleep furiously*

(2) *Furiously sleep ideas green colorless*

Tanto (1) quanto (2) sono dei nonsensi, ma qualunque parlante inglese riconoscerà che esse sono molto diverse quanto a buona formazione. Si può inoltre osservare che esistono frasi perfettamente dotate di senso che, tuttavia, qualunque parlante italiano avverte come non ben formate; ad esempio, la frase (3) appare comprensibile, al pari della frase (4), ma solo di quest'ultima possiamo dire che è grammaticalmente corretta:

(3) *Voglio di salutare Maria*

(4) *Ho voglia di salutare Maria*

La buona formazione, o grammaticalità, di una frase è quindi indipendente dalla sua immediata comprensibilità o dal fatto di poter essere effettivamente utilizzata per scopi comunicativi; viceversa, nell'uso concreto del linguaggio si possono incontrare frasi agrammaticali e, nonostante ciò, comprensibili.

Questa distinzione tra grammaticale ed accettabile, introdotta da Chomsky in una pubblicazione del 1965, si lega ad un'altra distinzione di importanza fondamentale: quella tra "competenza" ed "esecuzione". Con il primo di questi due termini si intende "la conoscenza che il parlante-ascoltatore ha della sua lingua", con il secondo di essi si fa riferimento all' "uso effettivo della lingua in situazioni concrete": *grammaticale* è quindi un concetto relativo alla competenza, mentre *accettabile* riguarda l'esecuzione. La comprensibilità, o sensatezza non è dunque condizione né necessaria (come mostra il contrasto tra (1) e (2)) né sufficiente (come mostra quello tra (3) e (4)) perché si possa parlare di buona formazione delle frasi e delle altre combinazioni di parole dal punto di vista sintattico: essa va quindi spiegata in base ad altri fattori, oggetto di verifica dell'analisi sintattica.

Nell'ambito dell'elaborazione del linguaggio naturale, con il termine "parsificazione" si intende l'operazione attraverso cui a partire da una frase in input si determina se essa appartiene al linguaggio L(G) di cui è nota una grammatica G e, in caso affermativo, si costruisce la sua rappresentazione sintattica; l'oggetto che esegue l'operazione viene chiamato "analizzatore sintattico" (*parser*).

5.4. L'analisi semantica

La determinazione del significato di una frase avviene in diverse fasi successive: nella prima ci si serve solo della struttura sintattica e dei significati delle parole, mentre nella seconda si tiene conto delle frasi precedenti già interpretate, di conoscenze sull'argomento di cui il testo tratta e di conoscenze generiche sul mondo. La prima fase prende il nome di "analisi semantica", e il suo risultato è la "forma logica" della frase.

La forma logica è una rappresentazione, in un linguaggio formale opportuno, del significato di una frase, per quanto è possibile determinarlo senza tenere conto né delle altre frasi del testo di cui è parte, né di conoscenze generiche sull'argomento trattato. La forma logica non dipende dalla struttura sintattica della frase: quella per "la bambina mangia la mela" deve essere indistinguibile da quella per "la mela è mangiata dalla bambina". Perché ciò sia possibile occorre disporre di una descrizione dei significati delle parole in cui si astragga dalle caratteristiche sintattiche. La strada comunemente seguita passa per la determinazione dei ruoli che i costituenti possono avere all'interno della frase. Così, per esempio, nel ruolo di "Agente" si troverà chi compie volontariamente un'azione ("la mela è mangiata *dalla bambina*"), nel ruolo di "Tema" la cosa o la persona su cui una azione si esercita ("la bambina mangia *la mela*"), oppure il soggetto di una azione involontaria ("*la mela* cadde a terra"). L'"Esperiente" è colui che è coinvolto in un qualsiasi processo psicologico ("*la bambina* vide la mela"), mentre il "Beneficiario" è colui a vantaggio del quale viene eseguita una azione ("la madre *le* raccolse la mela"). Altri ruoli di significato intuitivo sono lo "Strumento", la "Provenienza", la "Posizione" e la "Destinazione" (questi tre anche in senso traslato per indicare istanti o intervalli temporali, oppure trasferimenti di possesso), il "Co-agente" e il "Co-tema". Le informazioni sui ruoli assegnabili ad una parola e sulle modalità di attribuzione dei valori sulla base delle caratteristiche sintattiche devono essere contenute nelle voci del dizionario: esse costituiscono la "semantica lessicale".

Non esiste un unico modo per strutturare una forma logica. Una possibilità consiste nel servirsi di quadruple composte da:

un *operatore*;

un *nome*;

un *tipo*;

un *insieme di modificatori*.

Nel caso di forme logiche corrispondenti ad intere proposizioni l'“operatore” fornisce informazioni sul modo e sul tempo del verbo principale, mentre per quelle di sintagmi nominali specifica se il riferimento è definito o indefinito, e se l'entità descritta è di numero singolare o plurale. Il “nome” è un identificatore unico per l'entità cui si fa riferimento, sia che si tratti di una azione che di un oggetto, materiale o meno, mentre il “tipo” indica un concetto della base di conoscenza. L'“insieme dei modificatori”, infine, assume la forma di una lista i cui elementi specificano attribuzioni di ruoli nel caso di forme logiche relative ad azioni, oppure predicano caratteristiche particolari nel caso di forme logiche di sintagmi nominali. Di seguito sono riportati due esempi di forme logiche:

es: *la mela acerba*

(DEF/SING m1 MELA (ACERBA m1))

es: *Mario ha piantato un chiodo nel muro con il martello*

(PASSATO-PROSSIMO p1 AZIONE-PIANTARE

(AGENTE p1 (NOME m1 PERSONA “Mario”))

(TEMA p1 (INDEF/SING c1 CHIODO))

(DESTINAZIONE-LUOGO p1 (DEF/SING m2 MURO))

(STRUMENTO p1 (DEF/SING m3 MARTELLO))

La costruzione della forma logica procede attraverso l'applicazione di “regole semantiche”, che esprimono le modalità con cui è possibile determinare l'interpretazione parziale per un costituente a partire dalle interpretazioni parziali dei sottocostituenti. Queste regole possono ricorrere a “restrizioni di selezione” per specificare che possono essere applicate solo se le interpretazioni dei sottocostituenti soddisfano certe condizioni. Ad esempio, si può voler stabilire che il verbo “accendere” può essere interpretato con il significato di “far prendere fuoco” se ha come “tema” un corpo infiammabile; mentre può avere il significato di “mettere in funzione” se il “tema” è un dispositivo elettrico o meccanico.

Il risultato dell'analisi sarà un insieme di possibili interpretazioni semantiche per la struttura della frase in input, fra le quali sarà chiamato a discriminare un modulo che si serva di informazioni di contesto.

5.5. I diversi tipi di ambiguità

Nei linguaggi naturali l'ambiguità si può manifestare a diversi livelli. Già nel compiere l'analisi morfologica ci si può rendere conto che ad una stessa parola possono corrispondere diverse letture:

- può darsi che il suffisso sia ambiguo. "Mangia", decomposto in "mangi-" e "-a", può essere interpretato come una voce dell'indicativo oppure dell'imperativo a seconda del contesto;
- può darsi che anche la radice sia ambigua. "Amo" ("am-" e "-o") ha una interpretazione sia come nome che come verbo.
- ci possono essere più modi possibili per scomporre la parola in radice e affissi. "Militare" viene decomposto come "milit-" e "-are" nel caso del verbo e come "militar-" e "-e" nel caso dell'aggettivo.

Anche se una parola non è ambigua in nessuno dei sensi detti ora, è molto probabile che abbia più significati diversi, nessuno dei quali può essere scartato a questo stadio dell'analisi.

Quando si esegue l'analisi sintattica ci si imbatte nell'ambiguità che è inevitabilmente presente nelle grammatiche dei linguaggi naturali. A volte, le ambiguità sintattiche sono conseguenza di ambiguità lessicali, come in:

Una vecchia porta la sbarra

in cui "vecchia", "porta" e "sbarra" sono parole lessicalmente ambigue in quanto possono essere interpretate in vari modi: "vecchia" può essere un aggettivo (in questo caso si aggancia al nome "porta") o un nome; "porta" può essere interpretata come un nome o come un verbo, così come "sbarra". Queste ambiguità possono quindi dare luogo alle seguenti interpretazioni:

(Una vecchia (porta la sbarra))

(Una vecchia porta (la sbarra))

Nel primo caso il soggetto della frase, "la vecchia", compie l'azione di "portare la sbarra"; nel secondo il soggetto è "la vecchia porta" e il sintagma verbale è costituito da "la sbarra".

Lingue diverse presentano fonti di ambiguità differenti: in inglese è frequente che ci siano più nomi di seguito e non è facile stabilire quale di essi costituisca la "testa" del sintagma, e quali sono invece i modificatori. In italiano le sequenze di nomi sono assai rare, e l'uso più esteso delle preposizioni rende questo problema meno pressante. D'altra parte ci possono essere moltissimi modi per agganciare gli aggettivi e i sintagmi preposizionali tra loro e ai sintagmi

nominali (si ricordi che in questa fase ancora non ci si serve del significato delle parole). Ad esempio, la definizione di "gondola" recita:

"Caratteristica imbarcazione asimmetrica della laguna veneta, lunga, snella, a fondo piatto, con pettine decorativo a prora e copertura al centro".

Mentre, una volta stabilito che "caratteristica" in questo contesto è un aggettivo e non un nome, è facile interpretare correttamente "asimmetrica" e "della laguna veneta" come modificatori di "imbarcazione", non è affatto ovvio (per un sistema automatico di analisi del testo, si intende) che "lunga", "snella" ecc. non si riferiscano invece alla laguna veneta, oppure che "con pettine decorativo" non vada a modificare "fondo". Esiste un principio linguistico generale, detto "vincolo di non incrocio", che asserisce che le connessioni tra i sintagmi non devono incrociarsi. Ciò implica, ad esempio, che se si assume che "snella" si riferisca a "imbarcazione" allora non si può collegare "a fondo piatto" con "laguna veneta". Nonostante ciò le strutture possibili per la frase rimangono numerose.

La natura fortemente posizionale dell'inglese impedisce di invertire le posizioni del soggetto e del complemento oggetto di un verbo. In italiano, invece, questa inversione è consentita e viene impiegata per accentuare l'importanza del complemento oggetto. Nel parlato la distribuzione delle pause, e soprattutto l'intonazione, aiutano nella disambiguazione, ma in un testo scritto, se entrambi i sintagmi si accordano al verbo, non c'è modo di riconoscere il fenomeno senza tenere conto del significato delle parole. Ad esempio, la frase:

"La bambina mangia la mela"

può essere interpretata correttamente solo se si tiene in conto che il soggetto del verbo "mangiare" deve essere animato.

La molteplicità degli alberi sintattici per una medesima frase costituisce la fonte principale di complessità computazionale e, quindi, di inefficienza. Per questa ragione si tende ad eseguire in parallelo l'analisi sintattica e la costruzione della forma logica, in modo da servirsi dei vincoli di selezione e di altri controlli di tipo semantico per eliminare prima possibile le interpretazioni incoerenti. A dimostrazione di ciò, possiamo far ricorso a particolari frasi, la cui osservazione ci consente di trarre importanti osservazioni riguardo al problema dell'ambiguità dei linguaggi naturali; si tratta dei cosiddetti *garden paths*. Ad esempio, mentre una frase come:

Ho visto un'ala di farfalla col microscopio

(che è sintatticamente ambigua) è interpretata senza che l'ambiguità sia percepita (e cioè come se l'elaborazione fosse deterministica), in:

Dì al ragazzo che deve andare subito a casa di sbrigarsi

si ha un attimo di incertezza quando si incontra “di sbrigarsi”. Questo è probabilmente dovuto al fatto che l'ambiguità di “che deve andare subito a casa” era stata precedentemente risolta assumendo che fosse una frase dipendente (da “dire”), e che ciò che segue obbliga a modificare tale scelta. Questo fenomeno è un elemento in sostegno dell'ipotesi che tutte le diverse sorgenti di conoscenza siano utilizzate non appena possibile per risolvere le ambiguità. Anche in questo modo, tuttavia, esistono ambiguità che possono non venir eliminate. Ad esempio, esistono frasi inerentemente ambigue; nella frase:

Ho visto l'uomo con il binocolo

i vincoli semantici non aiutano a capire se il binocolo sia lo strumento con cui è avvenuta l'osservazione o un modo per caratterizzare la persona osservata. Questo non deve stupire, dal momento che la frase sarebbe ambigua anche per un essere umano: in casi simili solo la conoscenza del contesto può essere di aiuto.

In alcuni casi, possiamo osservare che frasi intuitivamente non ambigue non possono essere state disambiguate col semplice ausilio della semantica; prendiamo ad esempio:

I giudici condannarono i terroristi perché volevano salvare l'ordine costituito

e:

I giudici condannarono i terroristi perché volevano attentare all'ordine costituito

È solo la conoscenza generale del mondo (e del ruolo che hanno in esso giudici e terroristi) che permette di stabilire qual è il soggetto della frase dipendente.

5.6. Utilizzo dell'analisi sintattica nel contesto del progetto navigAbile

Lo strumento di analisi sintattica può essere utilmente impiegato nell'abito del progetto navigAbile lungo tre filoni:

1. Il parser può essere impiegato come strumento di **analisi della leggibilità** e complessità della frase. Generalmente la leggibilità di un testo viene valutata attraverso l'indice di Gulpease (indicatore di complessità sintattica); tale indicatore tiene conto soltanto della cardinalità di parole in ogni frase e del numero di lettere in ogni parola; tuttavia si può facilmente intuire che tale indicazione non è completa: si possono fare esempi di frasi complesse che presentano un alto indice di Gulpease (e quindi teoricamente di alta leggibilità) che non sono tuttavia costituite da una struttura

semplice; si tratta ad esempio di frasi che contengono un numero elevato di pronomi relativi, e quindi di frasi subordinate; tali frasi “impegnano” più di altre durante la lettura. Un esempio di indicatore di leggibilità potrebbe essere, ad esempio, la profondità dell’albero sintattico che viene prodotto dal parser; tale proposta si basa su un concetto intuitivo di complessità, secondo cui questa è proporzionale al numero di sintagmi annidati della frase che si sta considerando. Questa informazione sulla profondità dell’albero sintattico è disponibile solo nel caso in cui si utilizzino algoritmi di parificazione ricorsivi, e non è fornita dallo shallow parser che verrà presentato, in quanto produce una rappresentazione sintattica lineare (flat).

2. L’analisi sintattica permette la **disambiguazione** dei lemmi (necessaria per le parole che presentano significati multipli), e di conseguenza può essere impiegata come supporto alla traduzione dall’italiano verso linguaggi iconici. La disambiguazione è necessaria per la realizzazione di un traduttore automatico, ad esempio verso i linguaggi iconici come Bliss o PCS. Il parser analizza la frase e, nel caso una parola presenti possa appartenere a categorie grammaticali diverse, verifica che esista un’attribuzione della categoria grammaticale che dia luogo a una stringa che appartiene alla grammatica del parser. Diverse prove eseguite con grammatiche non contestuali, anche non necessariamente complesse, mostrano come un alto numero di ambiguità è rimosibile con l’analisi sintattica. Una volta eseguita la disambiguazione, nel caso si tratti di un lemma non appartenente al vocabolario di base, è inoltre possibile ricercare eventuali sinonimi che sono presenti nel vocabolario di base, e che sono quindi più semplici. La disambiguazione è fondamentale anche per il riconoscimento di alcune forme composte che in PCS vengono tradotte con un unico simbolo.
3. Le informazioni prodotte dall’analisi sintattica possono essere utilmente impiegate come supporto alla generazione automatica delle **keywords** di un testo. Normalmente le parole chiave di un testo possono essere generate automaticamente sulla base della frequenza di occorrenza nel testo stesso. Tuttavia la frequenza non è la sola informazione rilevante per il riconoscimento delle keywords: ad esempio articoli, ausiliari, congiunzioni, risultano statisticamente frequenti all’interno di testi comuni, ma evidentemente non sono rilevanti. Grazie all’output del parser si può unire l’informazione sulla frequenza di occorrenza a quella della tipologia di strutture

sintattiche (chunk) in cui la parola occorre; ad esempio, nei chunk nominali c'è spesso il soggetto della frase, e di conseguenza, una parola che ricorre frequentemente nei chunk nominali, è candidata a essere una parola chiave. Questo strumento di supporto redazionale potrebbe essere integrato nel sistema di Content Management a supporto dei redattori che aggiornano il portale con le nuove notizie.

4. Infine può essere impiegata come strumento di supporto all'**analisi delle concordanze**, nell'ambito della composizione delle frasi da parte degli utenti del sito. Si tratta del processo mediante il quale i verbi vengono coniugati con il tempo e la persona corretti, e i nomi e aggettivi accordano in genere e numero col soggetto. Ad esempio, dopo che l'utente ha selezionato i tre lemmi: "io" "mangiare" "panino", c'è un componente che realizza la coniugazione del verbo mangiare, riconoscendo che il soggetto della frase è "io", e coniugando coerentemente il verbo alla prima persona singolare.

Di seguito vengono riportati alcuni esempi di frasi con ambiguità che può essere eliminata con l'analisi sintattica:

"Marco sbarra la porta"

In questo esempio ci sono 3 tipi di ambiguità; la prima sta nella classificazione del lemma "sbarra" che può essere un sostantivo o una forma flessa del verbo "sbarrare". Anche la prima parola presenta ambiguità: "Marco" può essere inteso come un nome proprio di persona oppure come una forma flessa del verbo "marcare", in particolare corrispondente alla prima persona singolare, indicativo presente; infine "porta" può essere un sostantivo oppure come una forma flessa del verbo "portare".

"Il panino è stato mangiato"

In questo caso l'ambiguità sta nella classificazione del lemma "stato" che può essere un sostantivo o una forma flessa del verbo "essere".

Per giustificare l'affermazione che l'indice di Gulpease non è completo nella valutazione della leggibilità, riporto un esempio che mette in luce alcuni suoi limiti: un esempio di frase con alto indice di Gulpease (e quindi teoricamente molto leggibile) ma che è sintatticamente complessa è la seguente:

"Il bimbo, che aveva preso la mela dalla mamma di una sua amica, le portò una sorpresa"

5.7. Parser chunk-IT

Il parser con cui ho effettuato le prove per proporre una metodologia di valutazione avanzata della leggibilità di testi è stato realizzato dall'Istituto di Linguistica Computazionale di Pisa (Federici [33], Lenci [34] e Allegrini [35]).

L'algoritmo di parsificazione preso in considerazione si basa sulla seguente tecnica: vengono dapprima individuate tutte le **functional word** presenti in una frase; questo processo di identificazione serve per realizzare la suddivisione delle frasi in unità con una struttura sintattica ben specificata, dette **chunks**. Questa tecnica è quindi non ricorsiva, in contrapposizione alle tradizionali tecniche di analisi sintattica. L'output di uno shallow parser non è quindi il tradizionale albero sintattico, ma una struttura lineare (flat) con tutti i chunk allo stesso livello.

Esistono tre tipi di functional word: articoli, proposizioni e verbi ausiliari. L'idea di fondo è che questi componenti del linguaggio si prestano per una divisione naturale della frase in chunk, e sono generalmente non ambigui.

Tipicamente un chunk è costituito da una lemma principale, potential governor, costellato da un certo numero di preposizioni che si dispongono secondo un particolare pattern. Nel caso dell'italiano il potential governor è sempre l'ultima parola del chunk, e può coincidere o meno con l'head della frase. Il potential governor è l'elemento del chunk da cui dipendono gli altri elementi che compongono lo stesso chunk: ad esempio, nel chunk nominale *"la mia casa"*, il potential governor è la parola *"casa"*.

Si noti che un chunk è costituito esclusivamente da parole adiacenti; non sono ammessi chunk discontinui.

La struttura interna di un chunk è descrivibile con una semplice grammatica CFG (context-free grammar). Diversamente invece, le relazioni tra i vari chunk che costituiscono una frase, derivano dalle selezioni lessicali.

I risultati di prove su diversi corpus di testi (Abney, [31]) mostrano che un alto numero di ambiguità lessicali è rimovibile attraverso un parser chunk ben progettato.

Il vantaggio di utilizzare uno shallow parser risiede nel fatto che l'analisi viene arrestata a una granularità tale che ad un livello immediatamente successivo le ambiguità presenti non sono removibili. Un parser di questo tipo ha quindi un'intrinseca robustezza, che costituisce uno dei requisiti fondamentali dei moderni sistemi di analisi del linguaggio naturale.

Un ulteriore vantaggio sta nel fatto che questo tipo di parse è computazionalmente meno oneroso di un tradizionale parser ricorsivo.

Ovviamente il limite di questo approccio è la parziale specificazione di alcuni attributi sintattici; tuttavia queste informazioni sono necessarie solo per alcuni particolari tipi di elaborazioni, come l'Information Extraction (IE).

Esistono due macro categorie di chunk:

- ◆ **Phrase chunk:** sono i chunk che contengono un potential governor
- ◆ **Marker chunk:** contiene elementi che non possono essere potential governor quali congiunzioni coordinative e punteggiatura.

La seguente tabella mostra le diverse tipologie di chunk e i relativi potential governor.

Nome	Tipo	Potential gov	Esempio
ADJ_C	Chunk aggettivo	Aggettivo	"bello", "molto bello"
BE_C	Chunk predicativo	Aggettivo, participio passato	"è bello", "è caduto"
ADV_C	Chunk avverbiale	Avverbio	"sempre"
SUBORD_C	Chunk subordinativo	Congiunzione subordinativa	"quando", "dove"
N_C	Chunk nominale	Nome, pronome, verbo, aggettivo	"la mia casa", "io", "questo", "l'aver fatto", "il bello"
P_C	Chunk preposizionale	Nome, pronome, verbo, aggettivo	"di mio figlio", "di quello", "dell'aver fatto", "del bello"
FV_C	Chunk verbale	Verbo	"sono stati fatti", "rimangono"
G_C	Chunk gerundio	Verbo	"mangiando"
I_C	Chunk infinito	Verbo	"per andare", "per aver fatto"
PART_C	Chunk participio	Verbo	"finito"

Tabella 3: Tipologie di chunk per l'italiano

I pronomi relativi non compaiono in questa classificazione, e di conseguenza possono appartenere a diversi tipi di chunk.

5.8. Una metodologia per la valutazione della leggibilità di un testo

Di seguito viene proposta una metodologia per la valutazione della leggibilità di un testo utilizzando l'output prodotto dall'analisi sintattica compiuta mediante lo shallow parser appena introdotto; si tratta di un'evoluzione dell'indice di Gulpease, che mira a superarne i limiti evidenziati in precedenza.

L'idea è quella di attribuire un peso ai diversi tipi di chunk, in base al contributo di complessità che questi apportano alla frase; è intuitivo, ad esempio, che un chunk che introduce una subordinata influenza significativamente la complessità del periodo. Un discorso analogo si ha per i chunk preposizionali, in quanto fanno generalmente riferimento a concetti annidati.

Attualmente, nel progetto navigAbile sono presenti due diversi indicatori, per le due dimensioni di valutazione prese in considerazione finora: la leggibilità del testo viene valutata attraverso l'indice di Gulpease, mentre la comprensibilità attraverso l'indicatore di comprensibilità descritto in precedenza (I_{COMP}).

Normalmente, la classificazione di un testo, nel contesto di navigAbile, avviene attraverso la verifica dell'appartenenza dei due indicatori a opportuni intervalli. In altre parole, per classificare un testo come appartenente alla categoria Basic, sia l'indice di Gulpease, sia la percentuale di parole appartenente al VdB devono essere maggiori di una certa soglia.

Tuttavia un meccanismo di analisi a soglie fisse è troppo rigido, e quindi tipicamente ci si limita a verificare che gli indicatori abbiano un opportuno andamento (crescente) passando dalla versione *Normal* a quella *Basic*.

Si pone ora il problema di identificare un nuovo possibile indicatore di leggibilità che tenga conto delle informazioni sintattiche che vengono prodotte dal parser; l'idea è di arricchire il contenuto informativo dell'indice di Gulpease con le informazioni sulla struttura del periodo fornite dal parser; questo nuovo indicatore può essere ottenuto come media (eventualmente pesata) di una serie di features:

1. Numero medio di parole per frase (o in alternativa il numero medio di parole per chunk).
2. Numero medio di lettere per frase; questi primi due indicatori possono essere espressi attraverso l'indice di Gulpease.

3. Numero medio di chunk per frase. Si noti come questo indicatore ha un contenuto informativo differente rispetto al numero medio di parole per frase: il numero di chunk coincide con il numero di potential governor, e quindi con il numero di "unità fondamentali" di cui è composto il periodo.
4. Distribuzione delle tipologie di chunk. Questo indicatore parziale deve tener conto del fatto che esistono delle differenze a livello di leggibilità dei diversi chunk, dipendenti dal livello di annodamento dei concetti che questi inducono.

Questi tre indicatori (dei quattro citati, i primi due possono essere espressi con l'indice di Gulpease) devono assumere valori all'interno di uno stesso intervallo: si è stabilito che il valore massimo che possono assumere è 100, per coerenza con il range dei valori dell'indice di Gulpease. Al valore massimo corrisponde la massima leggibilità.

Possiamo quindi definire il nuovo indicatore, arricchendo il contenuto informativo dell'indice di Gulpease I_G attraverso l'informazione aggiuntiva derivante dall'analisi sintattica:

$$I_{LEGG} = a \cdot I_G + b \cdot I_C + c \cdot I_T$$

Dove I_C è un indicatore che dipende dal numero medio di chunk per frase, e I_T è un indicatore che dipende dalla tipologia di chunk presenti nel testo di riferimento.

L'indicatore dipendente dal numero medio di chunk per frase, può essere valutato con l'espressione:

$$I_C = \frac{100}{\left(\frac{N_C}{N_F} - 1 \right)}$$

Dove N_C è il numero di chunk totali presenti nel testo (esclusi quelli di punteggiatura).

Intuitivamente la sua espressione si giustifica con la seguente considerazione: in ogni frase, secondo la grammatica a cui si fa riferimento, deve contenere almeno due chunk (uno nominale e uno verbale). Nel caso in cui la frase contiene solamente due chunk, siamo di fronte alla complessità minima possibile, e difatti l'indicatore assume il valore massimo, cioè 100. Nel caso siano presenti più chunk, l'indicatore assume via via valori sempre più bassi.

Per quanto concerne l'indicatore che tiene conto della tipologia di chunk, si può pensare di attribuire un peso di leggibilità diverso ad ogni tipo di chunk, e ottenere l'indicatore totale come somma pesata della frequenza di occorrenza dei vari tipi di chunk che sono presenti nel

testo da valutare. Questi pesi possono essere ricavati su base statistica, facendo riferimento ad un corpus costituito da testi riconosciuti come molto semplici. In particolare, si è scelto di considerare un set di notizie tratte da “AltaFrequenza”, un settimanale contenente notizie semplificate, a cui collabora un team di linguisti, tra cui De Mauro.

Il set di notizie è stato analizzato con lo shallow parser e sono state ricavate delle statistiche sui circa 4000 chunk che costituiscono il testo di riferimento.

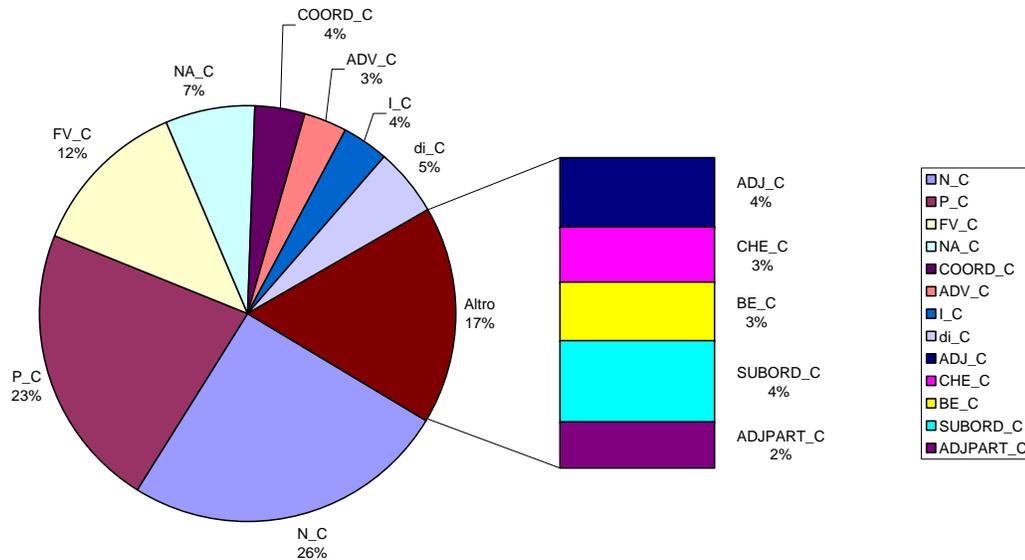


Figura 28: Distribuzione statistica delle tipologie di chunk nel corpus di AltaFrequenza

L'attribuzione dei pesi viene determinata in base alla frequenza statistica di ogni tipologia di chunk nel corpus di riferimento. Di conseguenza i chunk più frequenti avranno un peso di leggibilità minore e viceversa.

Con questa secondo modalità di attribuzione dei pesi, l'indicatore di complessità che tiene conto della tipologia dei chunk può essere valutato con la seguente relazione:

$$I_T = k \cdot \frac{\sum_{i=1}^T w_i \cdot N_i}{N_C}$$

Dove w_i è il peso del tipo di chunk i-esimo, mentre N_i è il numero di chunk del tipo i-esimo; k è invece una costante che serve per normalizzare a 100 il valore dell'indicatore I_T .

Come accennato in precedenza, al stima dei pesi w_i è stata calcolata in base alla frequenza di occorrenza di ogni tipo di chunk.

Tipo di chunk	Frequenza di occorrenza	Percentuale	Peso w_i
N_C	910	24.69%	0.2468
P_C	796	21.59%	0.2159
FV_C	445	12.07%	0.1207
NA_C	248	6,73%	0.0672
COORD_C	141	3.82%	0.0382
ADV_C	117	3.17%	0.0317
I_C	136	3.69%	0.0368
di_C	182	4.93%	0.0493
ADJ_C	140	3.80%	0.0379
CHE_C	107	2.90%	0.0290
BE_C	117	3.17%	0.0317
SUBORD_C	158	4.29%	0.0428
ADJPART_C	89	2.41%	0.0241
U_C	67	1.82%	0.0181
G_C	21	0.57%	0.0056
Altri chunk (unknown e PNA_C)	12	0.32%	0.0032

Tabella 4: Frequenze di occorrenze dei tipi di chunk nel corpus di AltaFrequenza e relativa attribuzione dei pesi

La costante di normalizzazione k viene stimata calcolando dapprima il valore dell'espressione

$$\frac{\sum_{i=1}^T w_i \cdot N_i}{N_C}$$

con la frase più semplice che è possibile costruire, che è costituita da due soli

chunk, uno nominale e uno verbale; successivamente si determina k in modo che per questo tipo di frasi l'indicatore di leggibilità valga 100; risulta $k = 544$.

Ricordiamo che PUNC_C sono i chunk che contengono punteggiatura (marker chunk). Questi tipi di chunk hanno una natura profondamente diversa rispetto agli altri tipi, e quindi, in questa trattazione, non verranno considerati per l'analisi di leggibilità.

E' evidente tuttavia che ad esempio il numero di virgole presenti in un testo ha un certo contributo informativo ai fini del calcolo della leggibilità; tipicamente infatti una virgola introduce una subordinata o un inciso.

Infine, per valutare l'indicatore complessivo di leggibilità: $I_{LEGG} = a \cdot I_G + b \cdot I_C + c \cdot I_T$, poniamo inizialmente $a = b = c = \frac{1}{3}$, che equivale a considerare la media aritmetica dei tre indicatori calcolati.

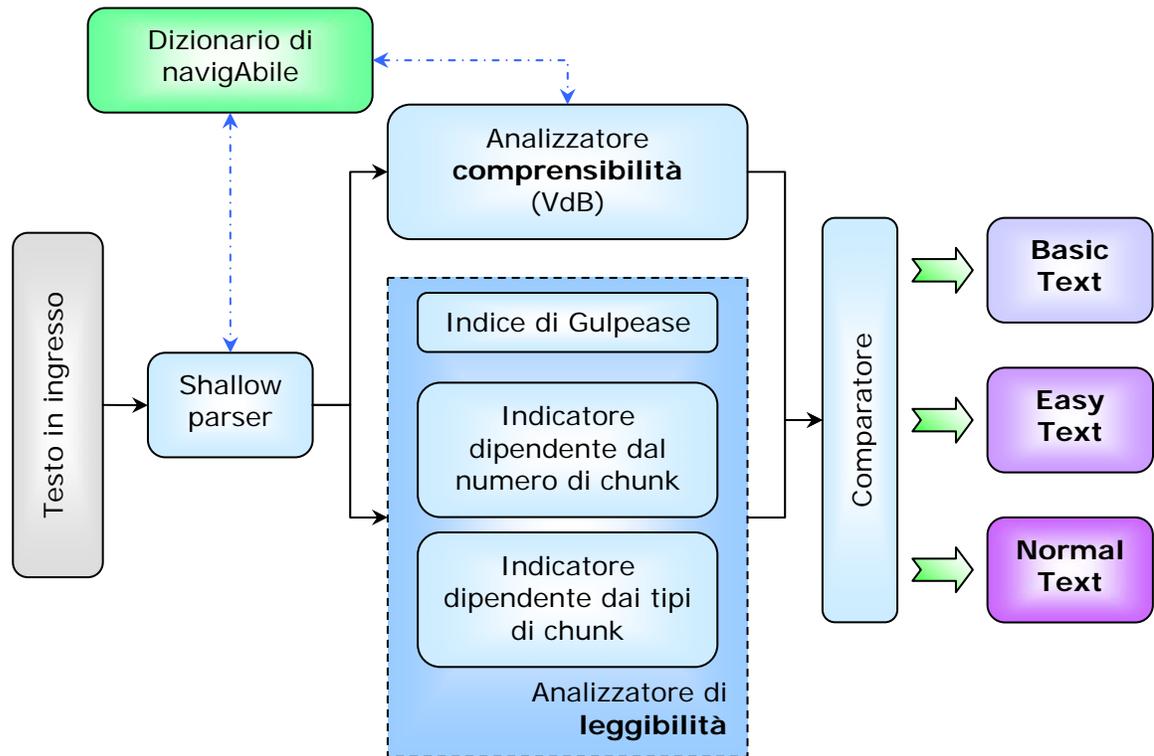


Figura 29: Schema del sistema di valutazione della comprensibilità e della leggibilità di un testo

Lo schema in figura mostra l'architettura del sistema di valutazione della comprensibilità e leggibilità che è stato proposto. Normalmente in ingresso vengono fornite le tre versioni semplificazioni di un certo testo, e vengono calcolati, per ognuna delle tre versioni, i due indicatori complessivi di leggibilità e comprensibilità. Successivamente il Comparatore confronta i due set di indicatori e verifica che sia la leggibilità che la comprensibilità siano crescenti passando dalla versione Normal a quella Basic, validando così la semplificazione che è stata eseguita manualmente partendo dal testo della notizia in versione Normal.

5.9. Risultati dell'analisi di leggibilità dei testi di prova

Per provare la validità della soluzione proposta, si è preso in considerazione un insieme di 16 notizie del sito, nelle tre versioni di semplificazione (Basic, Easy e Normal) e questo corpus è stato analizzato con lo shallow parser introdotto, e sono stati calcolati gli indicatori discussi in precedenza.

Di seguito viene riportata una tabella con delle statistiche sulle tipologie di chunk presenti nelle tre versioni delle notizie.

	Basic	Easy	Normal
Numero di chunk	709	1571	2571
N_C	160	309	517
FV_C	80	167	252
ADJ_C	20	43	83
COORD_C	42	95	151
P_C	99	232	395
I_C	29	50	77
BE_C	16	46	56
SUBORD_C	15	33	64
ADV_C	31	65	102
G_C	0	0	5
PART_C	0	0	0
ADJPART_C	8	23	42
NA_C	43	108	173
CHE_C	19	47	69
di_C	28	65	101
U_C	3	6	23
Altri chunk	2	8	13
PUNC_C	114	274	448

Tabella 5: Distribuzione delle tipologie di chunk nel testo di riferimento

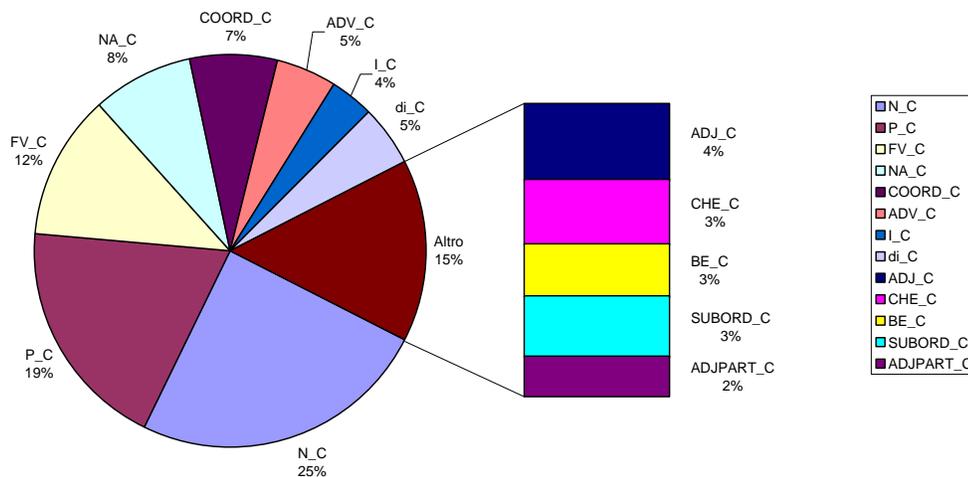


Figura 30: Distribuzione statistica delle tipologie di chunk nel corpus Normal

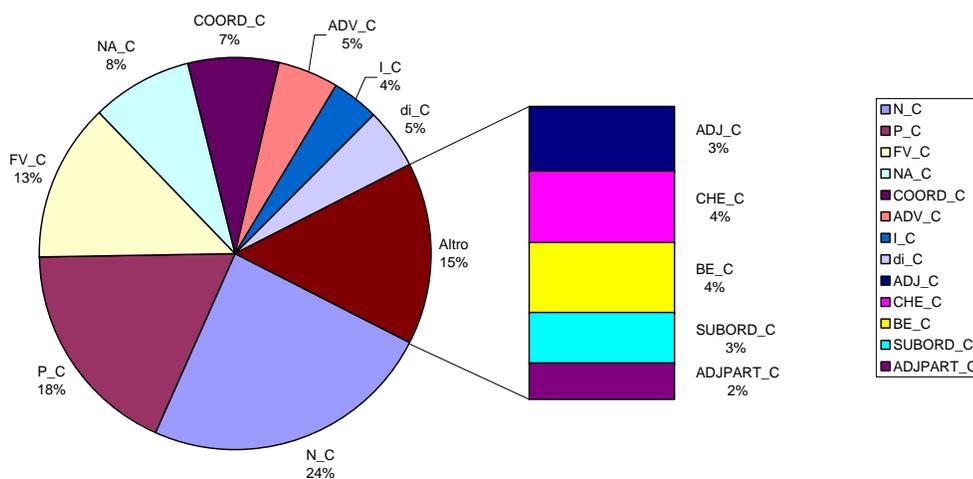


Figura 31: Distribuzione statistica delle tipologie di chunk nel corpus Easy

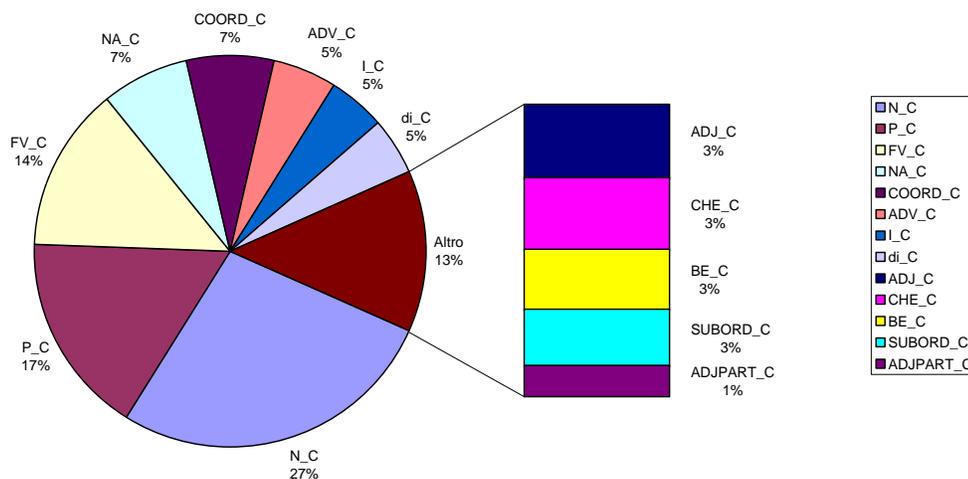


Figura 32: Distribuzione statistica delle tipologie di chunk nel corpus Basic

Un'analisi qualitativa su questi dati mette in evidenza i seguenti fatti:

- ◆ Passando da Basic a Normal si nota un aumento sensibile dei chunk preposizionali.
- ◆ Passando da Basic a Normal si nota una diminuzione dei chunk nominali.
- ◆ Passando da Basic a Normal il numero di chunk subordinativi e coordinativi aumenta sensibilmente.

Tutte queste considerazioni fanno intuire che, sulla base degli indicatori proposti, la leggibilità della versione Basic sia maggior di quella della versione Normal; questa ipotesi è confermata dai dati dalla seguente tabella.

	Basic	Easy	Normal
I_G	80.5	77.8	72.5
I_C	8.41	8.34	6.90
I_T	74.23	71.78	70.22
I_{LEGG}	54.3	53.0	51.3
I_{COMP}	80.3	79.1	76.8

Tabella 6: Valori degli indicatori di leggibilità e comprensibilità per i tre diversi livelli di semplificazione del testo

La tabella riassuntiva con i valori degli indici di leggibilità e comprensibilità evidenzia che:

- l'indicatore di leggibilità aumenta passando dalla versione Normal a quella Basic.
- l'indicatore di comprensibilità aumenta passando dalla versione Normal a quella Basic.

Con questo corpus di esempio, è stata quindi dimostrata la validità dello strumento proposto per valutare la qualità della semplificazione dei testi.

5.10. Analisi della traduzione dall'italiano verso i linguaggi iconici

Il metodo più semplice che risolve il problema della traduzione dall'italiano verso un linguaggio iconico è quello della traduzione parola per parola. I limiti di questo approccio sono evidenti: per tutti i lemmi che presentano significati multipli c'è ambiguità nella traduzione. Inoltre non è possibile riconoscere strutture grammaticali che potrebbero non essere presenti nel linguaggio di destinazione (come ad esempio i verbi ausiliari).

Intuitivamente si può ipotizzare che uno strumento di analisi sintattica della frase in italiano può garantire diversi potenziali miglioramenti:

1. risoluzione di alcune **ambiguità**; in particolare delle ambiguità che coinvolgono lemmi portano ad una struttura sintatticamente corretta solo quando appartengono ad una particolare categoria grammaticale.
2. l'output dell'analisi sintattica arricchisce il testo originale con delle informazioni che possono supportare l'eventuale processo di **semplificazione del testo**, ad esempio eliminando alcune frasi subordinate dell'albero sintattico.
3. le informazioni presenti nell'albero sintattico sono alla base della **traduzione** verso linguaggi che hanno una sintassi molto differente rispetto al linguaggio di partenza. Ad esempio si può capire a chi sono attribuiti gli aggettivi, e quindi si può eseguire l'eventuale inversione del sostantivo col verbo (come avviene nella traduzione dall'italiano verso l'inglese).

Essendo il PCS un linguaggio con una sintassi estremamente semplice, la necessità di introdurre un componente di analisi sintattica deve essere studiata.

Nel PCS non esistono i pronomi, quindi non è possibile costruire frasi subordinate; è quindi prevedibile che frasi scritte in PCS presentino intrinsecamente un'elevata leggibilità.

È prevedibile che con una grammatica non contestuale (CFG) abbastanza semplice si riesca a catturare una buona parte dei periodi che si possono formulare in PCS.

Nel linguaggio PCS non esiste inoltre un equivalente delle preposizioni, che sono presenti invece in alcune forme composte da più parole, come ad esempio "bustina di tè".

Viene proposto ora un metodo di valutazione della leggibilità, funzionale alla traduzione verso il linguaggio iconico PCS.

Bisogna quindi ridefinire l'indicatore che dipende dalla tipologia dei chunk.

Una possibile proposta per l'attribuzione dei pesi è la seguente: si attribuisce un peso maggiore ai chunk che non hanno un diretto equivalente in PCS. In questo modo si penalizza l'utilizzo di costrutti complessi, che vanno quindi tradotti con certo numero di costrutti più semplici.

Difatti non tutti i tipi di chunk introdotti sono direttamente traducibili in PCS; in altre parole solo alcune classi grammaticali presenti in italiano, sono presenti anche nel PCS.

In particolare in PCS sono presenti i seguenti tipi di chunk:

-
- ◆ ADJ_C
 - ◆ BE_C
 - ◆ ADV_C
 - ◆ N_C: solo se il POTGOV è un nome, un verbo, o un pronome personale
 - ◆ P_C: solo se il POTGOV è un nome, un verbo o un pronome personale
 - ◆ FV_C: non è possibile la forma passiva del verbo, né tempi diversi dal presente.

Mancano quindi le seguenti categorie:

- ◆ SUBORD_C
- ◆ I_C
- ◆ PART_C

Bisogna poi ricordare che nel PCS non sono presenti gli articoli, quindi nella traduzione da italiano verso PCS, tutti gli articoli non saranno tradotti. Nel processo di traduzione inverso, invece, bisogna introdurli in modo da rispettare le concordanze.

Un altro caso particolare è costituito dai verbi: in PCS non esistono modificatori per specificare il tempo del verbo e la persona (come succede, ad esempio, col Bliss), e quindi nella traduzione da PCS verso l'italiano, il verbo deve essere coniugato correttamente in modo da rispettare la concordanza con la persona che è soggetto del verbo.

L'indicatore che tiene conto della tipologia dei chunk viene ridefinito e può essere valutato con la relazione seguente:

$$I_T = \frac{h \cdot N_{PCS} + k \cdot N_{N-PCS}}{N_C}$$

Dove N_{PCS} è il numero di chunk che sono direttamente traducibili in PCS, N_{N-PCS} è il numero di chunk non direttamente traducibili in PCS, e N_C è il numero totale di chunk. Infine fissiamo i pesi: $h = 1, k = 0$. Con questi valori dei pesi l'indicatore assume valori nell'intervallo $0 \div 100$.

6. Conclusioni e Future Works

Possiamo riassumere il *principio di accessibilità* con la seguente citazione (Tommaso Tozzi):

"Costruire un'interfaccia il cui uso sia accessibile a tutti. Un'interfaccia che dunque non faccia uso di tecnologie il cui uso implichi attrezzature o software particolarmente costosi, complessi o con caratteristiche che ne impediscono l'uso anche solo ad alcuni. Garantire l'accesso alle informazioni di tipo pubblico senza alcun limite o discriminazione verso gruppi o individui. L'informazione vuole essere libera: appartiene a tutto il mondo, essa è prodotta da e per tutto il mondo e l'accesso all'informazione non deve più essere diritto esclusivo di un'élite o di un gruppo privilegiato."

Per l'utenza debole Internet costituisce una reale opportunità di integrazione sociale e quindi la tematica dell'accessibilità dei siti dovrebbe essere affrontata con un certo rigore. Il tema dell'accessibilità è quindi molto vasto e strutturato; è molto sentito poiché interessa un bacino di utenza considerevole.

Nonostante gli sforzi fatti in nel settore fino a questo momento, rimane sostanzialmente presente una relazione diretta tra accessibilità, da un lato, e costi e limitazioni grafiche e di interazione, dall'altro. In altre parole esistono una curva di trade-off che lega un aumento di dell'accessibilità di un sito all'aumento dei costi e dei compromessi necessari in termini di possibilità grafiche e di interazione.

Il principio basilare che sottende l'indipendenza degli strumenti e l'accessibilità è la separazione della forma dal contenuto; quando il significato di un documento è salvato separatamente da come deve apparire, l'indipendenza degli strumenti e l'accessibilità sono più facili da tutelare e questo lo si può ottenere con i fogli di stile (CSS), cioè una serie di istruzioni su come presentare o trasformare una pagina web: ad esempio si definisce la spaziatura e la dimensione del carattere, i colori, ecc.

Riassumendo possiamo individuare due strategie principali per affrontare il problema dell'accessibilità di portali web:

- ◆ Il primo approccio prevede l'impiego di versioni parallele. Il vantaggio è la possibilità di personalizzare ogni versione per il target di utenti a cui è rivolta; . Il limite di questo approccio è evidente: la gestione di siti paralleli è difficoltosa e onerosa. Gli aggiornamenti devono essere riportati su tutte le versioni parallele che sono state introdotte.

-
- ◆ Il secondo approccio, in accordo con il paradigma del "Design for All" prevede l'impiego di una sola versione che deve tener conto dei bisogni e delle esigenze di tutti gli utenti a cui è rivolto il portale. Evidentemente questa soluzione è il risultato di un compromesso tra accessibilità e evoluzione tecnologica e multimediale, e in un qualche modo limitare le tecniche di presentazione più avanzate.

La soluzione proposta con navigAbile si propone di unire i pregi dei due approcci, introducendo la funzione fondamentale della **profilazione e personalizzazione**, che permette una presentazione dei contenuti che si adatta alle **abilità** dell'utente. In questo senso sono pensati tutti gli strumenti di comunicazione multimodale presenti nell'architettura di navigAbile (linguaggi iconici, help audio...).

Il progetto navigAbile ha realizzato diversi prodotti:

1. **Servizi architetturali**: navigAbile è innanzitutto un'architettura orientata ai servizi (detta anche SOA o *Service Oriented Architecture*). I servizi architetturali sono quindi il primo prodotto del progetto e sono le "fondamenta" di navigAbile: Servizi di Interoperabilità, Profilazione, Filtro sulle pagine, Servizi di comunicazione, Servizi di supporto alla creazione di contenuti accessibili.
2. **Sito dedicato**: il sito dedicato di navigAbile (raggiungibile all'indirizzo www.navigabile.it) è nato per tre esigenze diverse: sottoporre a test i servizi architetturali, realizzando un primo esempio di applicazione che li mostrasse in azione; applicare i servizi navigAbile a un contesto semplice ma realistico di un sito di informazione; mettere a disposizione da subito agli utenti coinvolti nella sperimentazione alcuni servizi web accessibili, quali la bacheca o le news.

La realizzazione di questi prodotti ha portato anche alla definizione di un **metodo**, con cui sono stati identificati i cluster di utenti, definiti i requisiti, progettati, sottoposti a test, sperimentati e validati scientificamente i servizi di navigAbile.

Con l'introduzione di un'architettura orientata ai servizi viene proposto un modello di diffusione dei servizi per rendere accessibile un sito, che tende a ridurre i costi totali per l'accessibilità dei portali web, sfruttando la condivisione di servizi comuni a più siti.

Il lavoro svolto per l'analisi di comprensibilità e leggibilità ha proposto un nuovo metodo di valutazione di entrambi questi parametri di un testo.

L'indicatore di comprensibilità si basa sulla su statistiche di appartenenza delle varie parole del testo ai vari livelli del Vocabolario di Base di Tullio De Mauro e fornisce un'indicazione quantitativa sulla comprensibilità delle parole del testo, sulla base delle scelte lessicali dell'autore.

La metodologia introdotta per la valutazione della leggibilità mira ad arricchire il contenuto informativo dell'indice di Gulpease, attraverso l'impiego di informazioni derivanti dall'analisi sintattica, ottenute attraverso uno shallow parser. Questo approccio premia l'impiego di strutture sintattiche che sono statisticamente frequenti in un corpus di testi semplificati, presi come riferimento.

L'insieme di questi due indicatori consente di avere una comparazione quantitativa tra diversi testi e quindi in particolare, nell'ambito del progetto navigAibile, permette di valutare l'efficacia del processo di semplificazione a vari livelli (Normal, Easy, Basic) del testo di una stessa notizia.

Di seguito viene proposto un elenco di possibili miglioramenti e sviluppi futuri:

- ◆ Ipotesi di attribuzione dei pesi per il calcolo dell'indice di leggibilità complessivo, e relativa validazione, eventualmente effettuata con diverse categorie di utenti, in modo da testare l'effettiva comprensibilità dell'informazione che si vuole trasmettere.
- ◆ Ipotesi alternativa di attribuzione dei pesi dell'indicatore di leggibilità, basata sulle probabilità condizionate di concatenamento di chunk consecutivi: si tratta quindi di effettuare una stima di trovare una certa tipologia di chunk dati i precedenti, sulla base di un corpus di riferimento.
- ◆ Implementazione del componente architetturale per l'estrazione delle keywords, sviluppato sulla base delle informazioni rilevanti dell'analisi sintattica.
- ◆ Sfruttamento del parser per l'analisi delle concordanze: in particolare per realizzare la coniugazioni delle frasi che vengono composte con il relativo compositore.
- ◆ Word Predictor da integrare nel compositore frasi.
- ◆ Arricchimento dell'architettura con degli strumenti per la gestione del linguaggio Bliss, come ad esempio lo sviluppo di un componente per la traduzione da italiano a Bliss e viceversa.

7. Bibliografia

7.1. Riferimenti per il tema dell'accessibilità

1. "Tecnologie informatiche e utenza debole". Androni, Costa, Landoni. Editore Pirola. 2002. Pagine: 321.
2. "I disabili nella società dell'informazione. Norme e tecnologie". Ridolfi Pierluigi. Editore: Franco Angeli. 2000. Pagine: 192.
3. "Maximum Accessibility: Making Your Web Site More Usable for Everyone". John Slatin & Sharron Rush. Editore: Addison Wesley. 2002. Pagine: 633.
4. Web Content Accessibility Guidelines 1.0 e 2.0. W3C. <http://www.w3.org/WAI>
5. "Le web-barriere: l'accessibilità dei siti della Pubblica Amministrazione italiana". Gianluca Affinito.
6. "Tecnologie Informatiche e Utenza debole". Il Sole 24 Ore.
7. Libro Bianco: Tecnologie per la disabilità. Roberto Scanio.
8. "How People with Disabilities Use the Web". <http://www.w3.org/WAI/EO/Drafts/PWD-Use-Web/#usage>
9. Core Techniques for Web Content Accessibility Guidelines 1.0. In particolare il paragrafo su Multimedia Equivalents. <http://www.w3.org/TR/WCAG10-CORE-TECHS/#multimedia-equivalents>
10. Società specializzata in accessibilità (ha sviluppato il sito dell'unione italiana ciechi): <http://accessibilita.eraclito.it/>
11. Unione italiana ciechi: www.uiciechi.it
12. <http://www-3.ibm.com/able/>
13. <http://www.digitaldivide.org/>
14. <http://www.microsoft.com/enable/>
15. <http://www.webaccessibile.org>. Sito molto aggiornato per quanto concerne le normative italiane.
16. <http://www.cordis.lu/ist/so/einclusion/home.html>. Programma e-inclusion della comunità europea.
17. www.asphi.it. Associazione per lo Sviluppo di Progetti Informatici per gli Handicappati – ONLUS.
18. www.disabilityinfo.org. Portale governo USA su disabilità.

-
19. www.pubbliaccesso.it. Portale AIPA.
 20. http://www.cnipa.gov.it/site/it-IT/Le_Activit%c3%a0/Progetti/Accessibilit%c3%a0/
 21. www.superabile.it. Portale INAIL per disabili.
 22. www.italia.gov.it. Portale governo italiano.
 23. "User Requirements Document". World Wide Augmentative and Alternative – WWAAC.
<http://80.60.189.118/wwaac>
 24. "Understanding Accessibility". HiSoftware, R. B. Yonaitis.
 25. "Building Accessible Websites". Joe Clark.
 26. Siti simbolici e ad accesso facilitato per disabili cognitivi: www.peepo.com,
www.papunet.net, www.alta-frequenza.it
 27. Tesi di Laurea di Davide Danzi e Giorgio Gatti: "Bliss 2003". AA 2002-2003. Politecnico di Milano.

7.2. Riferimenti di linguistica

28. "Language, Volume 1 – An invitation to cognitive science". Daniel N. Osherson. Series Editor. 2000. Pagine: 429.
29. "Linguistica". Adrian Akmajian, Richard Demers, Robert Harnish, Farmer. Il Mulino. 1999. Pagine: 467.
30. "Syntactic Theory. A Formal Introduction". Ivan Sag, Thomas Wasow. CSLI Publications. 2001. Pagine: 480.
31. "Parsing By Chunks". Steven P. Abney. Pagine: 18.
32. "Part-of-Speech Tagging and Partial Parsing". Steven P. Abney. Pagine: 23.
33. "Chunking Italian: Linguistic and Task-oriented Evaluation". Stefano Federici, Simonetta Montemagni, Vito Pirrelli. Istituto di Linguistica Computazionale, CNR. Pagine: 8.
34. "Chunk-IT: an italian shallow parser for robust syntactic annotation". Alessandro Lenci, Simonetta Montemagni, Vito Pirrelli. Pagine: 35.
35. "POESIA: Public Open-source Environment for a Safer Internet Access". Paolo Allegrini, Nicoletta Calzolari, Simone Marchi, Simonetta Montemagni. Pagine: 30.

8. Elenco figure e tabelle

8.1. Indice delle figure

Figura 1: superabile.it in versione ad alta leggibilità	1-16
Figura 2: Pagina di personalizzazione del portale VisionConnection.....	1-16
Figura 3: Menù di personalizzazione del sito dell'Unione Italiana Ciechi	1-17
Figura 4: Homepage di peepo.com	1-18
Figura 5: Una pagina di contenuto del portale papunet.net.....	1-18
Figura 6: Il logo di navigAbile©: tutti i diritti riservati	2-22
Figura 7: Schema che evidenzia il Focus del progetto navigAbile.....	2-24
Figura 8: Schema dell'architettura di navigAbile	2-33
Figura 9: Sviluppo dell'albero di navigazione del prototipo iniziale	2-45
Figura 10: Homepage del prototipo in versione semplificata	2-46
Figura 11: Una pagina di indice del prototipo	2-47
Figura 12: Una pagina di contenuto del prototipo.....	2-47
Figura 13: Esempio di una voce del dizionario di navigAbile.....	2-48
Figura 14: Albero di navigazione del sito pilota navigabile.it	2-49
Figura 15: Homepage di navigAbile in versione standard	2-50
Figura 16: Pagina per la scelta del profilo	2-51
Figura 17: Homepage di navigAbile utilizzando il profilo 2	2-52
Figura 18: Homepage di navigAbile utilizzando il profilo 3	2-52
Figura 19: Pagina di una news in versione Normal (profilo 1)	2-53
Figura 20: Indice dei servizi di navigAbile (con profilo 2)	2-54
Figura 21: Pagina di indice delle sessioni del corso.....	2-55
Figura 22: Una pagina di contenuto del corso.....	2-56
Figura 23: Esempi di simboli Bliss.....	3-66
Figura 24: Esempi di procedure codificate con simboli PCS.....	3-67
Figura 25: Indice di Gulpease: scale dei valori.....	3-71
Figura 26: Schema delle tabelle della base di dati relativa dizionario di navigAbile	4-76
Figura 27: Il processo di analisi di un testo nell'elaborazione del linguaggio naturale.....	5-84

Figura 28: Distribuzione statistica delle tipologie di chunk nel corpus di AltaFrequenza	5-101
Figura 29: Schema del sistema di valutazione della comprensibilità e della leggibilità di un testo	5-103
Figura 30: Distribuzione statistica delle tipologie di chunk nel corpus Normal	5-105
Figura 31: Distribuzione statistica delle tipologie di chunk nel corpus Easy ...	5-105
Figura 32: Distribuzione statistica delle tipologie di chunk nel corpus Basic...	5-105

8.2. Indice delle tabelle

Tabella 1: Distribuzione dei diversi tipi di disabilità	1-8
Tabella 2: Soglie di leggibilità corrispondenti all'indice di Gulpease	3-71
Tabella 3: Tipologie di chunk per l'italiano	5-98
Tabella 4: Frequenze di occorrenze dei tipi di chunk nel corpus di AltaFrequenza e relativa attribuzione dei pesi	5-102
Tabella 5: Distribuzione delle tipologie di chunk nel testo di riferimento	5-104
Tabella 6: Valori degli indicatori di leggibilità e comprensibilità per i tre diversi livelli di semplificazione del testo	5-106

9. Ringraziamenti

Raggiungo questo importante traguardo dopo cinque anni di studio molto impegnativi ma altrettanto stimolanti, che ho vissuto con entusiasmo e che ricorderò sempre con piacere.

Desidero rivolgere un pensiero particolare ai miei genitori Pierino e Luisa, a mia sorella Elisa, a mio zio Ezio, e a tutta la mia famiglia, che mi ha sempre seguito con affetto, consigliandomi quando era necessario.

Rivolgo un sentito ringraziamento al project manager Giuliano Pozza, alla prof. Licia Sbattella e all'Ing. Antonio Bianchi per la loro disponibilità, competenza e l'aiuto nell'affrontare e risolvere le difficoltà che di volta in volta si sono presentate per realizzare questo lavoro.

Un doveroso ringraziamento va a tutte le persone del team di sviluppo di navigAble con cui ho collaborato durante lo stage in Accenture: Edoardo, Giovanni, Nando, Alexandre, Alessandro, Violetta, Elisabetta e i responsabili del progetto: Leopoldo, Bruno, Giorgio e Mauro. Li ringrazio tutti: durante questa esperienza ho potuto imparare molte cose, non solamente tecniche.

Voglio anche ringraziare i miei compagni di corso Davide Cervaro, Davide Coccia, Davide Chiarioni, Fabio Cristini, Elena Avanzi, Matteo Corno, Federico Fieschi, Lorenzo Bellini e Fabio Costa, con i quali ho condiviso ansie e gioie di cinque anni di Politecnico.

Ringraziamenti particolari vanno infine a Emanuele Varano, Federico Villa, Patrizia Agostino, Gigi d'Agostino, al prof. Sergio Rinaldi e al prof. Piero Baglione dell'ITIS G. Riva di Saronno. Da ognuno di loro posso dire di aver imparato qualcosa.

7 luglio 2004

Davide Ceriani

"La preoccupazione dell'uomo e del suo destino devono sempre costituire l'interesse principale di tutti gli sforzi tecnici. Non dimenticatelo mai in mezzo a tutti i vostri diagrammi ed alle vostre equazioni."

Albert Einstein (1879-1955)