Pubbliche funzioni e responsabilità 9

Saulle Panizza (a cura di)

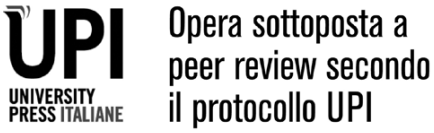
Profili attuali di qualità degli atti normativi e amministrativi

PISA
UNIVERSITY
PRESS

CIP

CIP a cura del Sistema bibliotecario dell'Università di Pisa

Pubbliche funzioni e responsabilità
Collana diretta da Saulle Panizza



Società con socio unico Università di Pisa
Capitale Sociale Euro 20.000,00 i.v. - Partita IVA 02047370503
Sede legale: Lungarno Pacinotti 43/44 - 56126, Pisa
Tel. + 39 050 2212056 Fax + 39 050 2212945
e-mail: press@unipi.it
www.pisauniversitypress.it

ISBN

Indice

Introduzione
Saulle Panizza ..... p. 5

Gli strumenti di riordino normativo, fra tecnica e politica
Fabio Pacini ..... » 9

La qualità della normazione con particolare riguardo
alla prospettiva del Parlamento
Francesco Dal Canto ..... » 31

Valutare l'impatto delle regole: un'occasione da non perdere
Francesco Sarpi ..... » 53

Alcune questioni sulla tecnica normativa, con particolare riferimento alle norme
sulla semplificazione contenute nella legge delega al Governo
in materia di riorganizzazione delle amministrazioni pubbliche
Saulle Panizza ..... » 65

Linguaggio discriminatorio e testi istituzionali: la questione
del genere grammaticale
Cecilia Robustelli ..... » 99

Le tecnologie linguistico-computazionali per la leggibilità
della comunicazione istituzionale
Dominique Brunato, Giulia Venturi ..... » 123

L'informatizzazione della legislazione vigente:
il progetto *Normattiva* e le banche dati pubbliche
Angela Di Carlo ..... » 163

Le applicazioni legimatiche
Pietro Mercatali ..... » 193

Adozione, diffusione e citazione della "Guida alla redazione degli atti amministrativi. Regole e suggerimenti": alcune esperienze
Francesco Romano » 211

La redazione degli atti normativi ed amministrativi nei contesti regionali, statali ed europei. Particolare riferimento alla scrittura degli atti amministrativi, ai problemi che si pongono e alle possibili soluzioni
Raffaele Libertini » 235

XmLeges-Editor: un ambiente completo per la redazione e conversione di testi normativi
Mariasole Rinaldi » 247

Le tecnologie linguistico-computazionali per la leggibilità della comunicazione istituzionale

Dominique Brunato, Giulia Venturi

Sommario: 1. La comunicazione tra istituzioni e cittadino. - 2. READ-IT: uno strumento automatico per l'analisi della leggibilità di un testo. - 2.1. Gli strumenti di Trattamento Automatico del Linguaggio. - 2.2. Il monitoraggio linguistico. - 2.3. Un esempio: la *Costituzione italiana* del 1947. - 3. READ-IT e un esempio di semplificazione legislativa. - 4. Conclusioni e sviluppi futuri. - Bibliografia

1. La comunicazione tra istituzioni e cittadino

La fase di rinnovamento che sta attraversando la Pubblica Amministrazione è legata a due principali fattori: da un lato, una presa di coscienza interna, che ha portato progressivamente a ridisegnare i termini stessi del rapporto tra l'amministrazione e il cittadino, dall'altro, la richiesta di adeguarsi alle dinamiche sempre più pervasive della comunicazione in rete, che fanno dell'accessibilità all'informazione uno dei principi ispiratori. Questa richiesta si fa ancora più pressante nell'era della digitalizzazione e degli Open Data: migliorare l'accesso all'informazione contenuta in grandi quantità soprattutto di testi scritti sta diventando infatti una questione fondamentale, come ricordato anche dalle Web Content Accessibility Guidelines[1] proposte dalla *Web Accessibility Initiative* e dalla *Europe 2020: Europe's growth strategy*[2], con la quale l'Unione europea ha posto tra i suoi obiettivi più ambiziosi quello di provvedere ad una "smart, sustainable and inclusive growth". Raggiungere l'obiettivo di una società inclusiva significa mettere in atto

[1] http://www.w3.org/TR/WCAG20/.

[2] http://ec.europa.eu/europe2020/pdf/europe_2020_explained.pdf.

azioni strategiche che permettano di rendere l'informazione facilmente accessibile ad un ampio ed eterogeneo gruppo di cittadini, compresi individui con difficoltà di lettura dovute ad un basso livello di scolarizzazione, al fatto che l'informazione non è veicolata nella lingua madre o ancora ad alcune specifiche disabilità linguistiche. È evidente che la comunicazione istituzionale, e soprattutto quella delle pubbliche amministrazioni chiamate a disciplinare aspetti della vita quotidiana dei cittadini, deve avere un ruolo di primo piano nella promozione di una società realmente inclusiva.

L'importanza di incentivare all'interno delle istituzioni pubbliche uno stile comunicativo improntato ai principi di chiarezza e semplicità linguistica diventa ancora più improrogabile se si considera qual è l'interlocutore di riferimento della Pubblica Amministrazione. Per sua natura, la comunicazione istituzionale non si rivolge a specialisti bensì al cittadino comune, che raramente è un esperto di materie amministrative. A ciò si aggiunge il quadro non troppo rassicurante che emerge dall'ultima indagine PIAAC (*Programme for the International Assessment of Adult Competencies*) svolta nel periodo 2011-2012 dall'OCSE (*Organizzazione per la Cooperazione e lo Sviluppo Economico*) in 24 paesi di Europa, America e Asia sulle competenze alfabetiche medie della popolazione adulta, lavoratori e non, di età compresa tra i 16 e i 65 anni[3]. In base a questa indagine, realizzata in Italia dall'ISFOL (*Istituto per lo sviluppo della formazione professionale dei lavoratori*), l'Italia si colloca all'ultimo posto per quanto riguarda le competenze alfabetiche (*literacy*) e al penultimo posto per le competenze matematiche (*numeracy*). I due tipi di competenze dovrebbero riflettere le conoscenze (i saperi) dell'individuo insieme ad alcune abilità di mettere in atto processi cognitivi di diversa complessità (comprensione testi, ragionamento, inferenze, deduzioni, calcolo, schematizzazione). L'importanza

[3] I paesi partecipanti sono stati: Australia, Austria, Belgio (Fiandre), Canada, Cipro, Danimarca, Estonia, Finlandia, Francia, Germania, Giappone, Irlanda, Italia, Norvegia, Paesi Bassi, Polonia, Regno Unito (Gran Bretagna e Irlanda del Nord), Repubblica ceca, Repubblica di Corea, Repubblica Slovacca, Russia, Spagna, Stati Uniti di America, Svezia. I risultati completi dell'indagine sono consultabili alla pagina http://www.isfol.it/pubblicazioni/research-paper/archivio-research-paper/le-competenze-per-vivere-e-lavorare-oggi.

dell'indagine per valutare il possesso di competenze chiave ritenute fondamentali per vivere e lavorare nelle attuali società è chiaramente evidente. Rispetto alle precedenti indagini OCSE la distanza rispetto agli altri paesi si è ridotta, tuttavia i risultati ottenuti non riflettono un miglioramento nella classifica del nostro paese rispetto.

Avere un quadro articolato del profilo del proprio destinatario è importante in ogni ambito della comunicazione istituzionale ed è una competenza fondamentale di chi si occupa, a vario titolo, di mettere in atto le pratiche suggerite dai diversi codici deontologici professionali oggi esistenti in Italia. Ad esempio, la *Carta dei doveri del giornalista degli uffici stampa* attribuisce al giornalista che opera nelle istituzioni il compito peculiare di operare "non solo per rendere riconoscibile l'Istituzione ai cittadini ma per farla da essi comprendere e rispettare"[4]. Analogamente, in ambito sanitario, è riconosciuto come presupposto del lavoro del comunicatore pubblico quello di "saper divulgare le informazioni corrette e allo stesso tempo chiare e comprensibili per i cittadini"[5]. Paradossalmente, l'attenzione rivolta alla questione della chiarezza linguistica nei codici deontologici della pubblica amministrazione è più altalenante. Se il *Codice di comportamento dei dipendenti delle pubbliche amministrazioni* del 2000, emanato dall'allora ministro Franco Bassanini, affermava chiaramente che "nella redazione dei testi scritti e in tutte le altre comunicazioni il dipendente adotta un linguaggio chiaro e comprensibile"[6], nella versione più recente del Codice, entrata in vigore il 19 giugno 2013, si è persa traccia di questa indicazione.

---

[4] La *Carta dei doveri del giornalista degli uffici stampa* è un documento approvato dal Gruppo Speciale Uffici Stampa dell'Ordine Nazionale dei Giornalisti il 26 febbraio 2002. È consultabile in rete all'indirizzo: http://www.odg.it/content/documento-cnog-26-febbraio-2002.

[5] Associazione Italiana della Comunicazione Pubblica e Istituzionale, *Documento di Indirizzo sulla Comunicazione Pubblica in Sanità*, maggio 2006 (disponibile al sito: http://www.marketingsociale.net/download/dcs.pdf).

[6] Presidenza del Consiglio dei Ministri, Dipartimento della Funzione Pubblica, decreto 28 ottobre 2000, Codice di Comportamento dei dipendenti delle pubbliche amministrazioni, pubblicato sulla G.U. n. 84 del 10 aprile 2001.

Questa decisione è solo una delle più recenti manifestazioni di una politica, non sempre coerente, che la pubblica amministrazione ha adottato nei confronti della questione della chiarezza linguistica. Sul piano dell'azione governativa, i risultati più significativi si sono ottenuti nel decennio compreso tra il 1990 e il 2000. La legge 241/90 ("Nuove norme in materia di procedimento amministrativo e di diritto di accesso ai documenti amministrativi") avvia di fatto il processo di rinnovamento interno alla Pubblica Amministrazione italiana e contribuisce a far maturare la consapevolezza che un'amministrazione trasparente è anche un'amministrazione che sa comunicare in maniera chiara, semplice ed efficace con il proprio destinatario. È da queste basi che prende avvio quel "moto di riforma del linguaggio amministrativo, che si prefigge di renderlo più chiaro e accessibile ai cittadini"[7] sulla scia dei risultati già ottenuti nel contesto anglosassone dal movimento *Plain Language*[8]. Tradizionalmente contraddistinto da caratteristiche di oscurità e complessità linguistica, tanto da essersi guadagnato appellativi negativi quali "*burocratese*", e ancora prima "*antilingua*"[9], il linguaggio della pubblica amministrativa italiana diventa così oggetto di numerose iniziative di semplificazione sia di ispirazione

[7] Come osserva Fortis (2005, p. 89), nonostante la legge 241/90 "non tratti esplicitamente del linguaggio, l'esigenza di una scrittura amministrativa più chiara costituisce un suo corollario ed è implicita nel suo spirito: consentire ai cittadini di accedere a documenti che comunque non riuscirebbe a comprendere sarebbe infatti un controsenso, che vanificherebbe, di fatto, tale diritto".

[8] Il *Plain Language* può essere definito come "the writing and setting out of essential information in a way that gives a co-operative, motivated person a good chance of understanding the document at first read, and in the same sense that the writer meant it to be understood" (Cutts 1998, p. 3). In molti paesi, sia europei che extra-europei, i principi di questo movimento sono stati recepiti concretamente dagli ordinamenti giuridici; un caso emblematico è quello della Svezia, dove è stata istituita una commissione presso il Ministero della Giustizia con il compito di verificare che le leggi siano scritte in *plain Swedish*. Per un approfondimento del *Plain Language*, si rimanda al contributo di Fortis (2003).

[9] È questo il termine coniato da Italo Calvino in un celebre articolo del 1965, in cui lo scrittore denuncia in maniera caricaturale, attraverso l'esempio di un verbale della polizia, come la lingua comune venga sottoposta a inutili complicazioni, prolissità e imprecisioni terminologiche nell'uso "burocratico".

governativa, sia coordinate da gruppi di lavoro provenienti da ambienti accademici e di ricerca. Tra il 1993 e il 2002, grazie all'impulso del Dipartimento della Funzione Pubblica, vengono pubblicati il *Codice di Stile ad uso delle Pubbliche Amministrazioni*, promosso dall'allora Ministro Sabino Cassese, il successivo – e più articolato – *Manuale di Stile. Strumenti per semplificare il Linguaggio delle Amministrazioni Pubbliche* (1997) e le direttive del Ministro della Funzione Pubblica dell'8 maggio 2002 ("*Direttiva sulla semplificazione del linguaggio dei testi amministrativi*") e del 24 ottobre 2005 ("*Direttiva sulla semplificazione del linguaggio delle pubbliche amministrazioni*"). In anni più recenti, un esplicito richiamo alla qualità dell'informazione diffusa dagli apparati istituzionali anche nei siti web ufficiali è contenuto nel decreto legislativo del 14 marzo 2013, n. 33, che individua, tra i requisiti di una comunicazione istituzionale di qualità, "la semplicità di consultazione, la comprensibilità, l'omogeneità, la facile accessibilità". Ancora, la legge 124/2015, all'articolo 16 del capo IV ("Deleghe per la semplificazione normativa"), rileva nella semplificazione del linguaggio una delle aree di intervento del programma più ampio di semplificazione normativa previsto dal Governo.

Anche nell'ambito della ricerca linguistica, il tema della chiarezza e della semplificazione della *lingua del diritto* nelle sue molteplici varietà ha storicamente riscosso ampio seguito. Ne sono nati non solo approfonditi manuali di stile e linee guida per la redazione dei testi istituzionali[10], ma anche metodi per il controllo della leggibilità dei testi che sfruttano tecnologie informatiche e statistiche per l'analisi testuale. In questo contesto, maggiore è stata anche la portata innovativa degli studi compiuti, soprattutto rispetto alla definizione di metodologie per valutare la qualità e l'accessibilità dei testi istituzionali basati sull'uso di indici automatici

[10] Si vedano, ad esempio, i manuali di Cortelazzo e Pellegrino (2003), Franceschini e Gigli (2003), Raso (2005) e, tra i contributi più recenti, la *Guida alla Redazione degli Atti Amministrativi. Regole e Suggerimenti*, realizzata da una équipe di ricercatori di formazione giuridica, linguistica e amministrativa, che si rifà alla Direttiva ministeriale del 2002 definendo, a sua volta, una serie di linee guida legate all'uso della lingua per produrre testi amministrativi chiari e precisi. La *Guida* è navigabile e scaricabile alla pagina http://www.pacto.it/content/view/416/48/.

per il calcolo della leggibilità. Sulla scia di quanto avveniva nel contesto anglosassone con il movimento del *Plain Language*, a partire dai primi anni '90, anche in Italia fiorirono numerosi studi sul calcolo della leggibilità di testi giuridici[11] che sfruttavano l'unico indice allora esistente: l'indice Gulpease (Lucisano e Piemontese 1988). Questo indice, analogamente agli indici di leggibilità tradizionali definiti per la lingua inglese, come ad esempio l'indice Flesch-Kincaid (Kincaid et al. 1975), approssima la complessità linguistica del testo attraverso due parametri molto semplici: la lunghezza della parola e la lunghezza della frase, considerati rispettivamente spie di complessità lessicale e sintattica in un testo. Sebbene basati su fattori di complessità linguistica piuttosto superficiali, che hanno una scarsa implicazione nei processi cognitivi di comprensione del testo, l'uso di questi indici è suggerito anche dai manuali di stile più recenti in materia di scrittura istituzionale[12].

Tuttavia, è soprattutto nell'ultimo decennio che, a livello internazionale, i metodi per la valutazione automatica della leggibilità dei testi hanno compiuto un salto di notevole qualità, tale da poter essere applicati con risultati promettenti anche nell'ambito della comunicazione istituzionale, come dimostrano lavori recenti compiuti per la lingua inglese (Katz e Bommarito 2014; Curtotti e McCreath 2013) e portoghese (Aluisio et al. 2010). Ad imprimere questa svolta è stato il concomitante sviluppo degli strumenti per il Trattamento Automatico del Linguaggio, che permettono oggi di ricostruire il profilo linguistico di un testo in maniera molto più articolata e affidabile del passato, andando a identificare uno spettro di caratteristiche di complessità linguistica ben più ampio di quello che gli indici di leggibilità tradizionali riusci-

[11] Per un approfondimento si possono consultare i numerosi lavori di Maria Emanuela Piemontese, tra cui: Piemontese 1990, pp. 225-246; 1996, pp. 123-193; 1999, pp. 269-292; 2000, pp. 103-117; 2001, pp. 119-130.

[12] Ad esempio, la *Guida alla redazione degli atti amministrativi* suggerisce l'uso di "applicazioni informatiche che facilitino l'applicazione delle regole (si pensi a sistemi di controllo della leggibilità degli atti e a editori specializzati simili a quelli già usati per gli atti normativi)" come mezzo per incentivare una più capillare diffusione dei modelli di scrittura proposti da parte dei dipendenti pubblici.

vano a computare automaticamente. Tali caratteristiche, rispecchiando le acquisizioni della linguistica formale e sperimentale sulla leggibilità e comprensibilità dei testi, hanno così contribuito a raffinare la definizione di complessità linguistica. Esse infatti corrispondono a una serie di fattori che, a vari livelli di descrizione linguistica, ostacolano la comprensione di una frase. Ad esempio, sul piano lessicale, un testo che presenta parole a bassa frequenza o astratte risulterà meno leggibile di un testo contenente parole più frequenti e concrete (Brown et al. 1987; Segui et al. 1982), soprattutto quando esso è rivolto a lettori non esperti. Rispetto alla struttura grammaticale, sono maggiormente costose, in termini di elaborazione cognitiva, le frasi in cui i costituenti legati da una relazione sintattica (ad esempio il soggetto e il verbo) sono distanti l'uno dall'altro (Gibson 1998; Frazier 1985) o disposti secondo un ordine non canonico[13] (Ferreira 2003). È significativo notare che questi, e altri tratti di complessità linguistica individuati in letteratura, sono tipicamente istanziati nel linguaggio burocratico, che infatti fa ampio uso di termini tecnici (anche quando non necessari, se non addirittura scorretti[14]), di parole obsolete e astratte, spesso frutto di nominalizzazioni, e, a livello sintattico, di incisi che interrompono la continuità della frase e strutture non canoniche, come ad esempio quelle in cui la frase subordinata precede la principale[15]. In quest'ottica, un indice di leggibilità linguisticamente motivato, ovvero in grado di cogliere gli effettivi luoghi di complessità del testo, può rappresentare un valido ausilio a supporto della redazione, ed

[13] In linguistica, l'ordine è un parametro che rende conto di come gli elementi (sintagmi o frasi) si dispongono in una struttura sintattica. Si parla di ordine canonico, o non marcato, per designare la disposizione naturale che gli elementi hanno in una determinata lingua e di ordine non canonico, o marcato, per identificare costruzioni che deviano da questo ordine. In lingue come l'italiano l'ordine di base vede il soggetto in prima posizione, seguito dal verbo e dal complemento oggetto (ordine "SVO"). Per un approfondimento si può consultare il link: http://www.treccani.it/enciclopedia/ordine-degli-elementi_(Enciclopedia_dell'Italiano)/.

[14] Si parla in questi casi di *tecnicismi collaterali* (Serianni 2005).

[15] Per una rassegna dei caratteri del linguaggio amministrativo si veda: Fortis 2005, pp. 57-89; Lubello 2014, pp. 45-61; Viale 2008, pp. 43-61.

eventuale riscrittura, di testi istituzionali ispirati ai principi di chiarezza, precisione ed efficacia comunicativa.

Il contributo che presentiamo si colloca in questo quadro di riferimento e si propone di illustrare lo stato dell'arte delle ricerche per la lingua italiana sulla valutazione automatica della leggibilità dei testi e a supporto della loro semplificazione. A questo proposito, sarà illustrato READ-IT (Dell'Orletta et al. 2011b), il primo e al momento unico strumento di valutazione della leggibilità oggi esistente per la lingua italiana basato su strumenti di Trattamento Automatico del Linguaggio e ispirato alla letteratura linguistica più recente sulla complessità linguistica. Attraverso gli esempi discussi nei paragrafi successivi, metteremo in evidenza come la metodologia alla base di READ-IT, che permette di segnalare in un testo aspetti di complessità lessicale e sintattica realmente implicati nei processi cognitivi di comprensione, possa rendere lo strumento un supporto efficace nelle diverse fasi del processo di *drafting* di testi istituzionali e incentivare l'adozione di pratiche di scrittura e revisione del testo coerenti con il modello linguistico della semplificazione.

2. READ-IT: uno strumento automatico per l'analisi della leggibilità di un testo

Gli ultimi anni hanno visto il progressivo affermarsi a livello internazionale del ricorso a tecnologie linguistico-computazionali per la misurazione automatica della leggibilità di un testo. A differenza dei metodi sino ad oggi adottati, come ad esempio la formula Flesch-Kincaid, utilizzata per la lingua inglese, o l'indice Gulpease per la lingua italiana, questa seconda generazione di strumenti di valutazione della leggibilità non fa affidamento unicamente su caratteristiche generali e formali del testo, quali la lunghezza della frase e la lunghezza delle parole. L'utilizzo di strumenti di annotazione linguistica automatica permette infatti di definire la leggibilità di un testo sulla base di parametri linguistici più complessi e che fino ad ora sembravano essere inattingibili se non attraverso un accurato lavoro manuale. Tali parametri spaziano tra i vari livelli di analisi linguistica e

sono rintracciati in modo automatico a partire dall'output del processo di annotazione automatica del testo.

Per quanto riguarda la lingua italiana, il primo e al momento unico strumento che si basa su questi presupposti è READ-IT (Dell'Orletta et al. 2011b) sviluppato dall'*Italian Natural Language Processing Laboratory* (ItaliaNLP Lab)[16] dell'Istituto di Linguistica Computazionale "Antonio Zampolli" (ILC) del CNR di Pisa[17] e concepito per fornire anche un supporto alla redazione semplificata di un testo attraverso l'identificazione dei suoi luoghi di complessità. READ-IT implementa un indice di leggibilità "avanzato" basato su analisi linguistica multi-livello del testo condotta con strumenti che rappresentano lo stato dell'arte per il trattamento automatico della lingua italiana. READ-IT, sulla base dei risultati del monitoraggio di una serie di caratteristiche linguistiche rintracciate in un corpus a partire dall'output di strumenti di annotazione linguistica automatica, permette di calcolare la leggibilità dei testi di cui il corpus è composto classificandoli come testi di *facile* o *difficile* lettura. La classificazione è realizzata da un classificatore statistico che associa i testi in input (linguisticamente annotati) a due classi di lettura definite a priori. Si tratta di classi formate da testi tratti dal corpus *Due Parole*[18], un giornale scritto con una lingua giornalistica volutamente semplificata per essere compresa da persone con un basso livello di scolarizzazione o con disabilità cognitive, considerati testi di *facile* lettura, e dal corpus *La Repubblica*, porzione del corpus CLIC-ILC (Marinelli et al. 2003), considerati testi di *difficile* lettura. L'appartenenza ad una delle due classi è stabilita sulla base del grado di similarità tra la distribuzione di alcune delle caratteristiche linguistiche monitorate. Ad esempio, testi con valori di ricchezza lessicale, lunghezza delle relazioni di dipendenza, lunghezza di sequenze di complementi preposizionali modificatori di teste nominali, ecc. più vicini ai valori di monitoraggio linguistico di *Due Parole* sono

[16] www.italianlp.it.

[17] Una demo on-line di READ-IT è disponibile alla pagina http://www.italianlp.it/demo/read-it/.

[18] http://www.dueparole.it/default.asp.

classificati come testi di facile lettura rispetto a testi che mostrano valori più simili a quelli di *La Repubblica*.

Un tratto caratterizzante di READ-IT, innovativo rispetto alla letteratura internazionale in materia, consiste in una valutazione della leggibilità articolata su due livelli: il documento e la singola frase. La valutazione rispetto alla frase rappresenta un'importante novità dell'approccio sottostante a READ-IT: attraverso l'identificazione dei luoghi di complessità del testo (individuati a livello della singola frase) che necessitano di revisione e semplificazione, lo strumento risulta essere un utile ausilio per la semplificazione del testo (Dell'Orletta et al. 2014b).

Ampiamente sperimentato su diverse tipologie di testi (Dell'Orletta et al. 2014a), READ-IT è stato sino ad oggi utilizzato per valutare l'efficacia comunicativa di testi in diverse tipologie di comunicazione: quella tra insegnante-studente, per fornire un supporto all'insegnante nella personalizzazione della sua azione formativa; operatore di call center-utente, per fornire un supporto alla redazione dei testi usati nei call center migliorando i processi di comunicazione con l'utente; medico-paziente, per assistere la redazione di consensi informati semplici e leggibili. In questo contributo l'intento è quello di dimostrare come READ-IT possa essere usato con successo per calcolare il livello di leggibilità di testi giuridici e per valutare l'efficacia della comunicazione legislatore e/o amministratore-cittadino, allo scopo di semplificare e migliorare i processi di comunicazione tra istituzioni e cittadini.

In quanto segue, saranno prima descritti gli strumenti di Trattamento Automatico del Linguaggio su cui si basa READ-IT (§ 2.1); sarà poi introdotta la metodologia di monitoraggio linguistico alla base del calcolo della leggibilità (§ 2.2); nel paragrafo 2.3 sarà infine presentato un esempio di analisi della leggibilità di un testo realizzata con READ-IT.

2.1. Gli strumenti di Trattamento Automatico del Linguaggio

Gli strumenti di Trattamento Automatico del Linguaggio, operando in successione, permettono di rendere progressivamente esplicita l'informazione linguistica contenuta in un testo. Per ogni livello di descrizione linguistica uno specifico componente di analisi identifica in modo auto-

matico la struttura del testo, utilizzando come input il risultato prodotto dal componente precedente. L'identificazione della struttura linguistica del testo, o annotazione, avviene tipicamente in modo incrementale, attraverso analisi linguistiche a livelli di complessità crescente: "tokenizzazione", ovvero segmentazione del testo in parole ortografiche (o *token*); analisi morfo-sintattica e lemmatizzazione del testo *tokenizzato*; analisi della struttura sintattica della frase in termini di relazioni di dipendenza.

La tabella 1 mostra un esempio del risultato del processo incrementale di annotazione linguistica del seguente periodo estratto da una direttiva comunitaria in materia ambientale:

Le disposizioni di cui alla presente lettera si applicano anche nei confronti degli altri organi tenuti all'adozione di strumenti urbanistici.

Innanzitutto, il periodo è stato individuato grazie alla fase di segmentazione del testo in periodi. Durante la successiva fase di tokenizzazione, all'interno del periodo sono stati riconosciuti i token corrispondenti alle singole forme (colonna *Forma*), identificate univocamente da un numero progressivo (colonna *Id*). La fase di disambiguazione morfo-sintattica ha permesso di associare ad ogni token individuato *i)* la corretta categoria morfo-sintattica (colonna *CPoS* e *FPoS*)[19] che il token ha nel contesto specifico, *ii)* i relativi tratti morfologici (colonna *Tratti morfologici*) e *iii)* il lemma corrispondente (colonna *Lemma*). Ad esempio, la forma 'disposizioni' (Id=2) viene ricondotta al lemma 'disposizione', viene annotata con la categoria sostantivo (S) e viene inoltre riconosciuto che si tratta di una forma plurale (num=p) e femminile (gen=f).

Il risultato dell'annotazione sintattica riportato nelle colonne *Testa sintattica* e *Relazione* della tabella 1 permette inoltre di stabilire che, ad esempio, il sostantivo 'disposizioni' è il soggetto (subj) del verbo 'applicano', il quale costituisce la testa sintattica della relazione. Questa informa-

[19] Per ogni token viene riconosciuta la categoria morfo-sintattica generale (CPoS) e eventuali sottocategorie (FPoS). Ad esempio, alla forma (token) 'alla' viene associata la categoria preposizione (E) e viene ulteriormente specificato che si tratta di una preposizione articolata (EA). Allo stesso modo, il token '.' viene annotato come un segno di punteggiatura (F) di fine periodo (FS).

Tabella 1. Un esempio di annotazione linguistica.

		Lemmatizzazione	Annotazione morfo-sintattica			Annotazione sintattica	
Id	**Forma**	**Lemma**	**CPoS**	**FPoS**	**Tratti morfologici**	**Testa sintattica**	**Relazione**
1	Le	il	R	RD	num=p|gen=f	2	det
2	disposizioni	disposizione	S	S	num=p|gen=f	9	subj
3	di	di	E	E	-	5	comp
4	cui	cui	P	PR	num=n|gen=n	3	prep
5	alla	a	E	EA	num=s|gen=f	2	mod_rel
6	presente	presente	A	A	num=n|gen=n	7	mod
7	lettera	lettera	S	S	num=s|gen=f	5	prep
8	si	si	P	PC	num=n|per=3|gen=n	9	clit
9	applicano	applicare	V	V	num=p|per=3|mod=i|ten=p	0	ROOT
10	anche	anche	B	B	-	9	mod
11	nei	in	E	EA	num=p|gen=m	9	comp
12	confronti	confronto	S	S	num=p|gen=m	11	mod
13	degli	di	E	EA	num=p|gen=m	12	prep
14	altri	altro	A	A	num=p|gen=m	15	mod
15	organi	organo	S	S	num=p|gen=m	13	prep
16	tenuti	tenere	V	V	num=p|mod=p|gen=m	15	mod
17	all’	a	A	EA	num=s|gen=n	16	comp
18	adozione	adozione	S	S	num=s|gen=f	17	prep
19	di	di	E	A	-	18	comp
20	strumenti	strumento	S	S	num=p|gen=m	19	prep
21	urbanistici	urbanistico	A	A	num=p|gen=m	20	mod
22	.	.	F	FS	-	9	punc

zione è riportata nella colonna *Testa* dove è infatti segnalato che la testa sintattica del dipendente 'disposizioni' ha Id=9, l'Id cioè del token 'applicano'. In questo caso 'applicano' ha testa sintattica 0 dal momento che rappresenta il verbo della frase principale, radice (root) dell'albero sintattico dell'intero periodo. La fase di annotazione sintattica a dipendenze permette dunque di fornire una descrizione esplicita dell'intero albero sintattico del periodo analizzato, sotto forma di relazioni di dipendenza che legano i token che lo compongono. L'informazione può inoltre essere graficamente visualizzata, come mostra la figura 1 che riporta la struttura sintattica della frase annotata, rappresentata come una serie di nodi lessicali (i singoli token), messi in collegamento da archi di dipendenza a loro volta etichettati con il nome del tipo di relazione di dipendenza (gli archi e le etichette graficamente rappresentati).

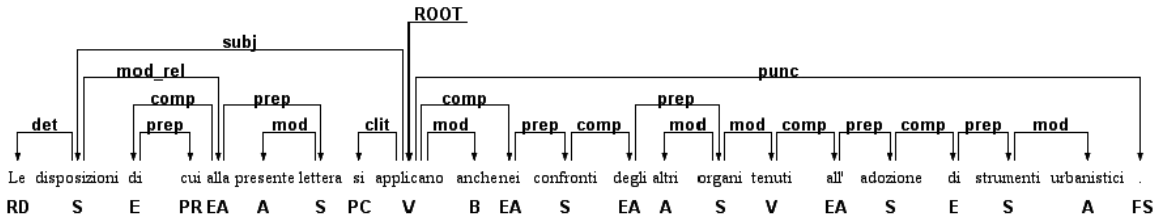


Figura 1. Un esempio di rappresentazione grafica dell'annotazione sintattica a dipendenze.

In questo studio è stata utilizzata *LinguA* (*Linguistic Annotation pipeline*), una catena di strumenti statistici di Trattamento Automatico del Linguaggio sviluppati in modo congiunto dall'Istituto di Linguistica Computazionale "Antonio Zampolli" (ILC) del CNR di Pisa e dall'Università di Pisa[20]. Tali strumenti rappresentano lo stato dell'arte per la lingua italiana essendo risultati gli strumenti più precisi e affidabili nell'ambito della campagna di valutazione di strumenti per l'analisi automatica dell'italiano, EVALITA-2009[21]. In particolare, il modulo di annotazione morfo-sintattica ha dimostrato un'accuratezza del 96,34% (Dell'Or-

[20] Una demo di *LinguA* è disponibile alla pagina http://linguistic-annotation-tool.italianlp.it/.

[21] http://www.evalita.it/2009.

letta 2009)[22] nell'identificazione simultanea della categoria grammaticale e dei tratti morfologici associati. Per quanto riguarda l'analisi a dipendenze, il modulo di annotazione sintattica a dipendenze realizzato dal parser *DeSR* (Attardi et al. 2009) raggiunge livelli di LAS[23] e UAS[24] in linea con lo stato dell'arte dell'analisi a dipendenze, pari a 83,38% e 87,71% rispettivamente.

Come sottolineato per la prima volta da Gildea (2001), gli strumenti di Trattamento Automatico del Linguaggio hanno una drastica diminuzione di accuratezza quando sono impiegati nell'analisi di tipologie di testi rappresentativi di un dominio diverso da quello sui quali gli strumenti sono stati sviluppati. Si tratta della questione nota come *Domain Adaptation*, attività di ricerca volta a definire metodologie di adattamento degli strumenti all'analisi di testi che appartengono a un dominio diverso da quello rispetto al quale sono stati sviluppati. Il dominio giuridico non rappresenta un'eccezione, come recentemente dimostrato dal "Domain Adaptation Track" (Dell'Orletta et al., 2013a) di EVALITA-2012[25] e dal "First Shared Task on Dependency Parsing of Legal Text" (Dell'Orletta et al. 2012) dell'edizione 2012 del workshop "Semantic Processing of Legal Texts" (SPLeT-2012)[26], dove sono state messe a punto una serie di metodologie per adattare strumenti di Trattamento Automatico del Linguaggio sviluppati per l'analisi di testi giornalistici, considerati rappresentativi della lingua comune, ai testi giuridici sia italiani sia inglesi. Particolare interesse è stato dedicato all'analisi di quali aspetti in-

[22] L'accuratezza è calcolata come il rapporto tra il numero di tokens classificati correttamente e il numero totale di tokens analizzati.

[23] LAS (*Labelled Attachment Score*) è una metrica che indica la proporzione di parole del testo che hanno ricevuto un'assegnazione corretta per quanto riguarda sia la testa sintattica sia la dipendenza che le lega.

[24] UAS (*Unlabelled Attachment Score*) è una metrica che indica la proporzione di parole del testo che hanno ricevuto un'assegnazione corretta per quanto riguarda l'identificazione della testa sintattica.

[25] http://www.italianlp.it/software/evalita-2011-domain-adaptation-for-dependency-parsing/.

[26] http://www.italianlp.it/software/first-shared-task-on-dependency-parsing-of-legal-texts-at-splet-2012/.

fluiscono di più sul grado di accuratezza dell'annotazione sintattica. Una tale attenzione è legata al fatto che questo livello costituisce il punto di partenza per numerose applicazioni pratiche, quali ad esempio l'estrazione automatica di informazione, la traduzione automatica, il *Question Answering*, ecc. Per superare questo problema, gli strumenti statistici di Trattamento Automatico del Linguaggio usati in questo lavoro sono stati adattati unendo due *training* corpora rappresentativi di due diversi domini: la treebank ISST-TANL (Dell'Orletta et al. 2012), composta da articoli di giornale considerati rappresentativi della lingua comune, e il corpus TEMIS (Venturi 2012a), una collezione di testi legislativi e amministrativi italiani annotata fino a livello sintattico, rappresentativi del dominio giuridico. Questo ha permesso di mantenere l'accuratezza degli strumenti allo stato dell'arte.

2.2. Il monitoraggio linguistico

Come precedentemente introdotto nel Paragrafo 2, il calcolo della leggibilità operato da READ-IT si basa sui risultati del monitoraggio di una serie di caratteristiche linguistiche rintracciate in un corpus a partire dall'output dei diversi livelli di annotazione linguistica: lemmatizzazione, annotazione morfo-sintattica e annotazione sintattica a dipendenze. Grazie a tale metodologia di monitoraggio, il profilo linguistico di un testo è ricostruito sulla base della distribuzione di tratti linguistici che spaziano tra diversi livelli di descrizione linguistica: lessicale, morfo-sintattico e sintattico. La tabella 2 riporta alcuni dei tratti monitorati illustrati nei paragrafi che seguono.

Prendendo le mosse da una più generale metodologia di monitoraggio della lingua italiana nelle sue varietà diamesiche, diastratiche, diafasiche introdotta per la prima volta da Dell'Orletta e Montemagni (2012) e Montemagni (2013) ai quali si rinvia per una descrizione dettagliata, tale metodo è stato sperimentato su varie tipologie di testi, come ad esempio le produzioni scritte e i materiali didattici offerti nella scuola primaria e secondaria allo scopo di monitorare le competenze linguistiche di apprendenti l'italiano come L2 (Dell'Orletta et al. 2011a). Per quanto riguarda il dominio giuridico, un confronto tra il

Tabella 2. Alcune delle caratteristiche considerate in fase di monitoraggio linguistico da READ-IT.		
Tipo di caratteristica	**Livello di annotazione linguistica**	**Caratteristica**
Di base	Divisione in frasi	Lunghezza media dei periodi e delle parole
Lessicale	Lemmatizzazione e annotazione morfo-sintattica	Percentuale di lemmi appartenenti al *Vocabolario di Base* del *Grande dizionario italiano dell'uso* (De Mauro 2000)
		Distribuzione dei lemmi rispetto ai repertori di uso (Fondamentale, Alto uso, Alta disponibilità)
Morfo-sintattico	Annotazione morfo-sintattica	Distribuzione delle categorie morfo-sintattiche
		Densità lessicale
Sintattico	Annotazione sintattica a dipendenze	Distribuzione dei vari tipi di relazioni di dipendenza sintattica (es. soggetto, oggetto)
		Arità verbale
		Caratteristiche relative alla struttura dell'albero sintattico analizzato: altezza media dell'intero albero sintattico; lunghezza media della più lunga relazione di dipendenza sintattica
		Caratteristiche relative all'uso della subordinazione: distribuzione di frasi principali vs. subordinate; lunghezza media di sequenze consecutive di subordinate
		Caratteristiche relative alla modificazione nominale: lunghezza media dei complementi preposizionali dipendenti in sequenza da un nome

profilo linguistico di diversi tipi di testi (atti legislativi, atti amministrativi e sentenze) rappresentativi di diverse varietà della lingua del diritto è descritto in Venturi (2012b).

2.3. La Costituzione italiana *del 1948 analizzata con READ-IT*

Come esempio di output di READ-IT, abbiamo scelto di riportare i risultati di un esperimento condotto sulla *Costituzione italiana* nella sua versione originaria del 1948. La scelta è motivata dall'intenzione di verificare l'uso della metodologia di valutazione della leggibilità qui proposta su un tipo di testo legislativo a lungo studiato sia da linguisti, sia da giuristi, sia da esperti di informatica giuridica. Tali lavori hanno dimostrato come la nostra Costituzione sia caratterizzata da una prosa che si distingue per una "scorrevolezza e relativa facilità di lettura della nostra Carta fondamentale in confronto alla grande maggioranza dei testi normativi italiani" (Mortara Garavelli 2001) a dimostrazione di uno "straordinario impegno dei *Costituenti*" e di un "non comune impegno linguistico" (De Mauro 2006).

Come illustrato nella figura 2, l'interfaccia di READ-IT permette di copiare e incollare nella scheda *Testo da analizzare* il testo di cui si intende calcolare il livello di leggibilità.

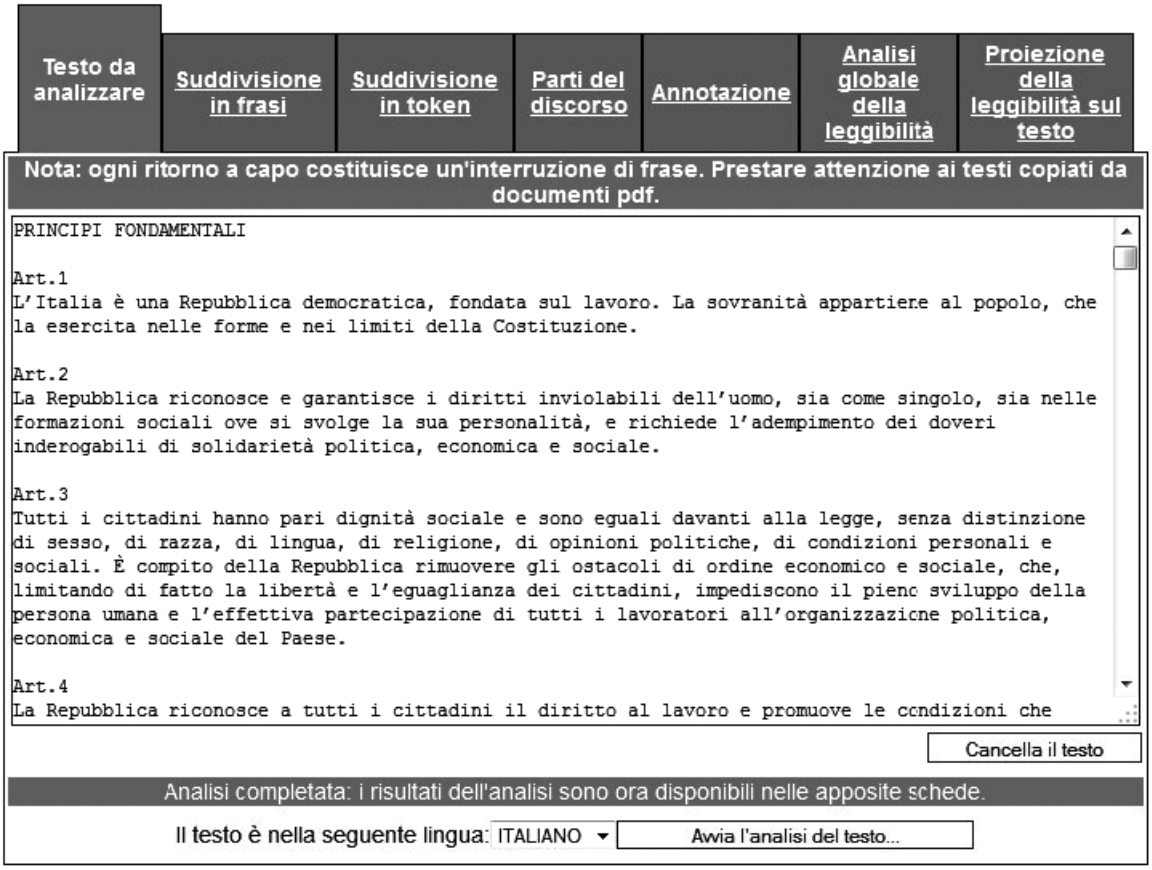


Figura 2. Il testo della *Costituzione italiana* del 1947 da analizzare.

Una volta che gli strumenti di Trattamento Automatico del Linguaggio hanno annotato linguisticamente il testo in input, è possibile visualizzare il risultato del calcolo della leggibilità nella scheda *Analisi globale della leggibilità*, come si può vedere nella figura 3[27]. Oltre al calcolo del valore di Gulpease, READ-IT conduce la valutazione globale della leggibilità del testo rispetto a quattro diversi indici calcolati sulla base di quattro diverse configurazioni di caratteristiche del testo:

– *BASE*: in questo modello, le caratteristiche considerate sono quelle usate nelle misure tradizionali della leggibilità di un testo (ovvero la lunghezza della frase e la lunghezza delle parole).
– *LESSICALE*: questo modello si focalizza sulle caratteristiche lessicali del testo (ovvero la composizione del vocabolario e la sua ricchezza lessicale).
– *SINTATTICO*: questo modello si basa su informazione di tipo grammaticale, ovvero sulla combinazione di tratti morfo-sintattici e sintattici.
– *GLOBALE*: si tratta di un modello basato sulla combinazione di tutti i tratti considerati dagli altri modelli.

Per ciascun modello, la percentuale esprime il livello di difficoltà, ovvero si riferisce alla probabilità di appartenenza del testo in esame alla classe dei testi di *difficile* leggibilità[28]: la barra a fianco esprime visivamente questo valore, dove il rosso rappresenta la probabilità di appartenenza alla classe dei testi difficili e il verde a quelli di facile lettura.

Nel caso specifico, la *Costituzione italiana* ha un valore di difficoltà di lettura del 98,3% dato dal modello *GLOBALE* e risulta dunque un testo di difficile lettura. A differenza del punteggio di leggibilità dato dall'indice Gulpease, pari a 53,9[29] (riportato nell'interfaccia di READ-IT), READ-IT fornisce un punteggio diverso a seconda del modello di calcolo della leggibilità considerato. Rispetto, ad esempio, al modello *BASE*,

[27] La versione a colori dell'immagine è disponibile alla pagina http://www.italianlp.it/wp-content/uploads/downloads/figure__readit/Figura-3.png.

[28] I punteggi di leggibilità di READ-IT vanno dunque da 0 a 100: più il valore percentuale è basso, più il testo in esame è semplice.

[29] In base alla scala di leggibilità di Gulpease, un testo con punteggio pari a 53,9 è un testo di difficile lettura per chi ha la licenza media. Rispetto a questo indice, la *Costituzione* risulta dunque un testo di media difficoltà di lettura.

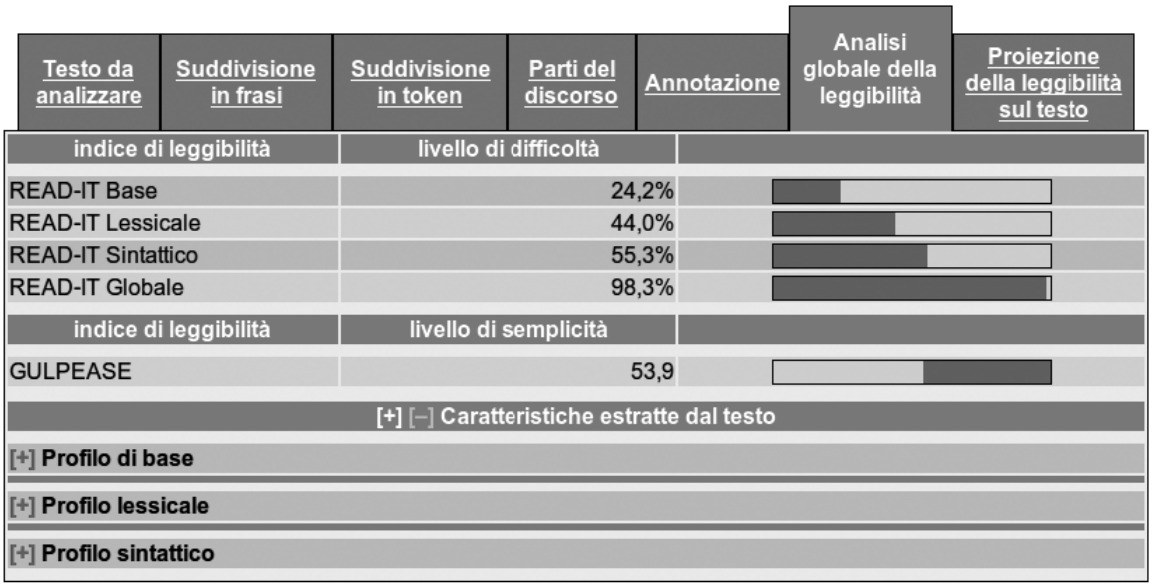


Figura 3. Risultato del calcolo globale della leggibilità della *Costituzione italiana*.

la *Costituzione* si rivela semplice, con un livello di difficoltà del 24,2%. Analogamente, anche sulla base del modello *SINTATTICO*, che tiene in considerazione le caratteristiche sintattiche, il testo risulta meno complesso (55,3%) rispetto al modello globale basato sull'intero insieme di caratteristiche linguistiche.

Un'analisi completa di tali differenze può essere condotta tenendo in considerazione le caratteristiche catturate da READ-IT in fase di monitoraggio linguistico del testo. Come si può vedere nella figura 3, la sezione dell'interfaccia *Caratteristiche estratte dal testo* riporta i risultati del monitoraggio di un sottoinsieme (selezionato come significativo) delle caratteristiche linguistiche utilizzate da READ-IT nella misurazione della leggibilità. Se consideriamo, ad esempio, i tratti considerati nella ricostruzione del *Profilo di base* del testo analizzato (vedi figura 4[30]), notiamo che la *Costituzione* contiene frasi con una lunghezza media pari a circa 16 token (16,1) per frase, una lunghezza che si avvicina di più a quella dei testi di facile lettura (che contengono frasi con una lunghezza media pari

[30] La versione a colori dell'immagine è disponibile alla pagina http://www.italianlp.it/wp-content/uploads/downloads/figure__readit/Figura-4.png.

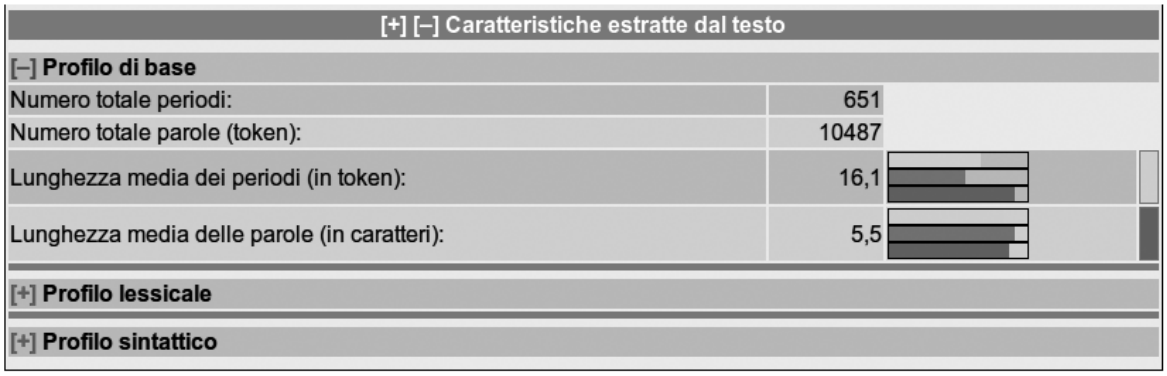


Figura 4. Caratteristiche linguistiche del *Profilo di base* della *Costituzione italiana*.

a 19 token) che non a quella dei testi di difficile lettura (con lunghezza media di frasi pari a 27 token)[31]. Nell'intera sezione, per ogni caratteristica riportata, oltre al valore numerico, viene fornita una rappresentazione grafica che mette a confronto il dato relativo al testo oggetto dell'analisi (corrispondente alla barra azzurra) con la corrispondente informazione rilevata nei corpora di riferimento di facile (barra verde) e difficile (barra rossa) lettura. Il rettangolino a fianco fornisce una classificazione semantica del dato rilevato in relazione al testo oggetto dell'analisi.

Il punteggio di leggibilità ottenuto sulla base sia del modello *LESSICALE* (44,0%), che tiene in considerazione le informazioni lessicali rintracciate all'interno del testo considerato, sia del modello *SINTATTICO* (55,3%) ci restituisce una *Costituzione* non eccessivamente difficile né dal punto di vista lessicale né sintattico. Il dato è in linea con quanto fatto osservare da Tullio De Mauro riguardo al "non comune impegno linguistico" (De Mauro 2006) dei padri costituenti verso la redazione di un testo leggibile e comprensibile per la maggior parte di cittadini. Se ci concentriamo infatti sui risultati del monitoraggio linguistico del *Profilo lessicale* (vedi figura 5[32]) notiamo che, ad esempio, la distribuzione di lemmi che appartengono al *Vocabolario di Base* (VdB) della *Costituzione* è

[31] Per visualizzare nell'interfaccia web i valori dei testi di facile e difficile lettura di riferimento è sufficiente passare con il cursore sulla barra verde o rossa.

[32] La versione a colori dell'immagine è disponibile alla pagina http://www.italianlp.it/wp-content/uploads/downloads/figure__readit/Figura-5.png.

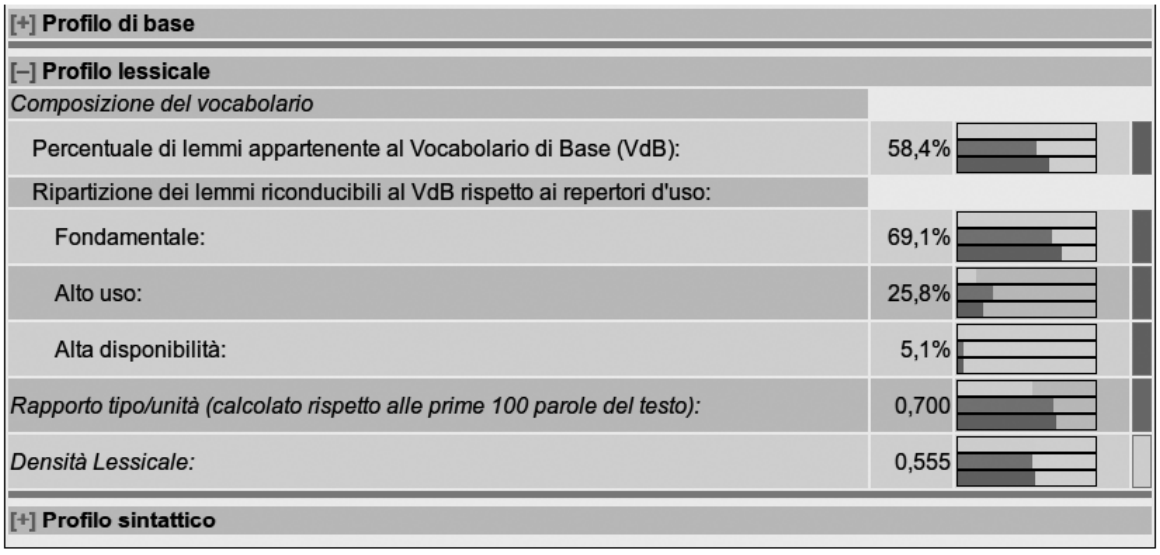
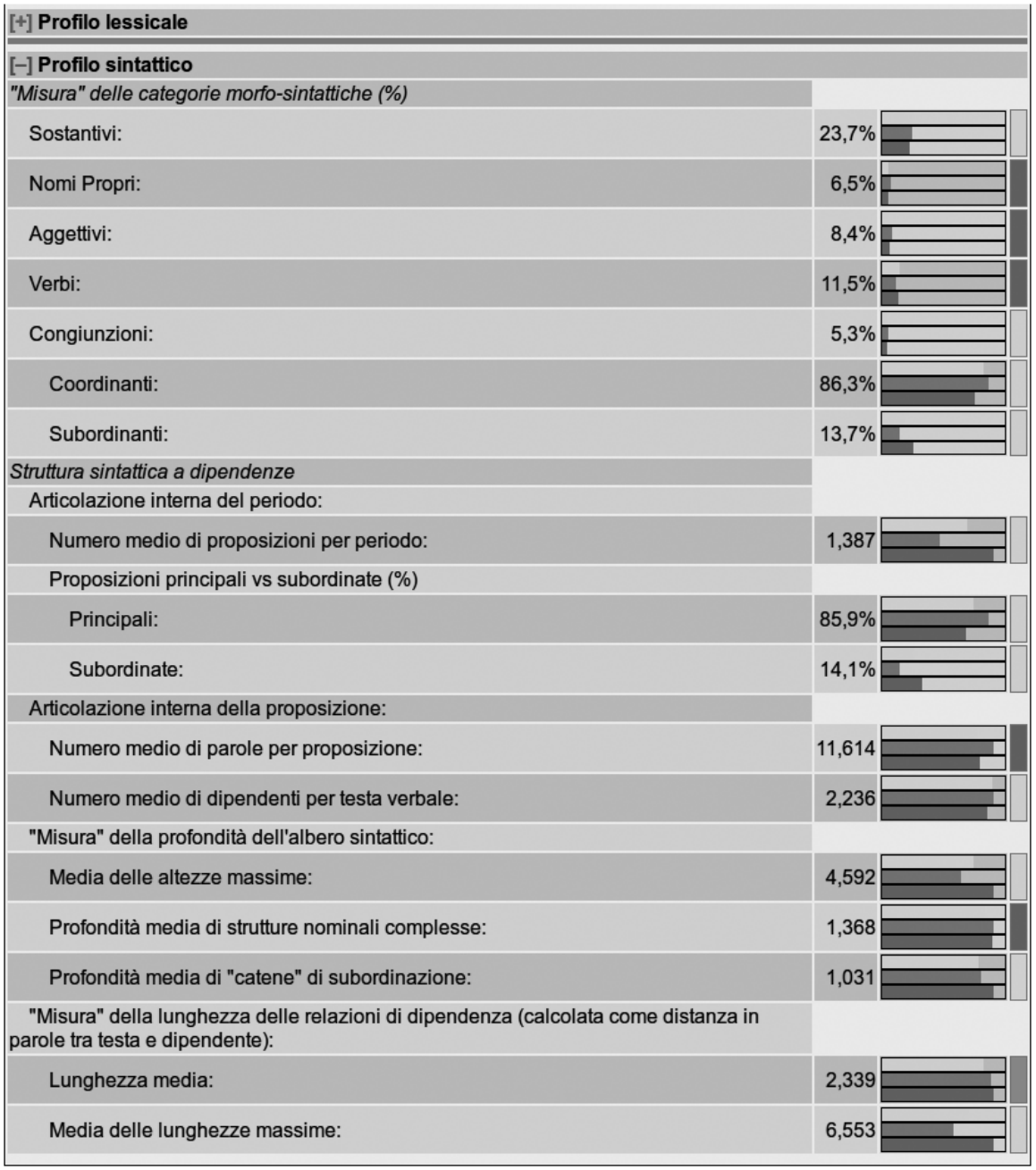


Figura 5. Caratteristiche linguistiche del *Profilo lessicale* della *Costituzione italiana*.

pari al 58,4%. O ancora più in particolare possiamo osservare che i lemmi della *Costituzione* che appartengono al repertorio del *Lessico Fondamentale* sono pari quasi al 70%; ciò è indicativo di un testo pensato per essere leggibile ad un ampio pubblico di lettori.

Se andiamo poi ad analizzare i tratti tenuti in considerazione nella ricostruzione del *Profilo sintattico* (vedi figura 6[33]), notiamo che la *Costituzione* ha caratteristiche sintattiche più simili ai testi di semplice lettura, indicatori di semplicità che influiscono poi sul calcolo della leggibilità rispetto al modello *SINTATTICO*. Per alcune caratteristiche sintattiche il testo mostra addirittura un profilo di facilità di lettura ancora maggiore di quello dei testi qui considerati di facile lettura. È il caso, ad esempio, della media delle altezze massime degli alberi sintattici (*Media delle altezze massime* nella figura 6), calcolata come il numero di relazioni di dipendenza sintattica consecutive; mentre i testi di facile lettura (la barra verde di riferimento) hanno valori medi pari a circa 5 relazioni di dipendenza consecutive per frase, la *Costituzione* ha un valore pari a circa 4,6 relazioni. Le stesse differenze si ritrovano anche per quanto riguarda la lunghez-

[33] La versione a colori dell'immagine è disponibile alla pagina http://www.italianlp.it/wp-content/uploads/downloads/figure__readit/Figura-6.png.



[+] **Profilo lessicale**	
[–] **Profilo sintattico**	
"Misura" delle categorie morfo-sintattiche (%)	
Sostantivi:	23,7%
Nomi Propri:	6,5%
Aggettivi:	8,4%
Verbi:	11,5%
Congiunzioni:	5,3%
Coordinanti:	86,3%
Subordinanti:	13,7%
Struttura sintattica a dipendenze	
Articolazione interna del periodo:	
Numero medio di proposizioni per periodo:	1,387
Proposizioni principali vs subordinate (%)	
Principali:	85,9%
Subordinate:	14,1%
Articolazione interna della proposizione:	
Numero medio di parole per proposizione:	11,614
Numero medio di dipendenti per testa verbale:	2,236
"Misura" della profondità dell'albero sintattico:	
Media delle altezze massime:	4,592
Profondità media di strutture nominali complesse:	1,368
Profondità media di "catene" di subordinazione:	1,031
"Misura" della lunghezza delle relazioni di dipendenza (calcolata come distanza in parole tra testa e dipendente):	
Lunghezza media:	2,339
Media delle lunghezze massime:	6,553

Figura 6. Caratteristiche linguistiche del *Profilo sintattico* dei primi 12 articoli della Costituzione.

za media delle relazioni di dipendenza sintattica (*Media delle lunghezze massime*), calcolata come la distanza tra la testa e il dipendente (in token); mentre i testi di facile lettura hanno lunghezze medie pari a circa 8 token, la *Costituzione* contiene frasi con relazioni di dipendenza lunghe in media 6,5 token.

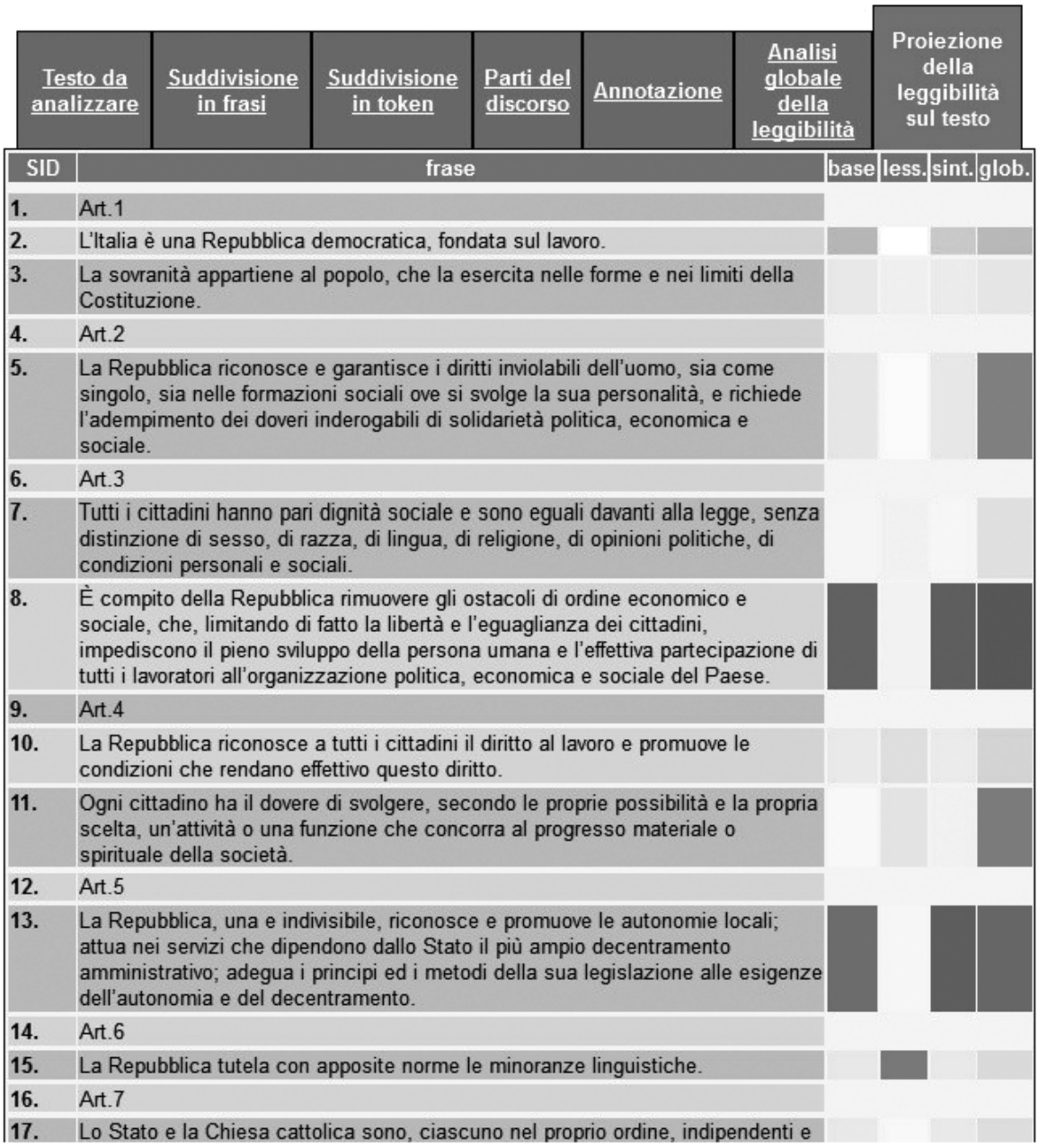


Figura 7. Leggibilità delle singole frasi contenute nella *Costituzione italiana*.

Come precedentemente fatto notare, un tratto caratterizzante READ-IT, innovativo rispetto alla letteratura internazionale in materia, consiste in una valutazione della leggibilità articolata su due livelli: il documento e la singola frase. La valutazione rispetto alla frase è stata esplicitamente concepita per fornire un supporto al redattore del testo e guidarlo nel processo di revisione e semplificazione. I risultati di questo livello più granulare di calcolo della leggibilità sono contenuti nella scheda *Proiezione*

della leggibilità sul testo dove è possibile identificare le frasi che necessitano di revisione. Come si può vedere nella figura 7[34], per ogni frase viene riportato il livello di difficoltà base (*base*), lessicale (*less.*), sintattico (*sint.*) e globale (*glob.*) in colonne distinte, livello calcolato dai corrispondenti modelli di analisi della leggibilità. Il livello di difficoltà è rappresentato cromaticamente mediante colori che vanno dal verde (frase leggibile) al rosso (frase particolarmente difficile): il rosso, così come sfumature giallo-arancioni, marcano frasi che necessitano di revisione.

Mentre dunque una frase come la terza è mediamente semplice, con un punteggio di leggibilità pari a 54,4%, l'ottava frase è decisamente più complessa con un punteggio di 96,5%[35]. Questo è un segnale per il redattore che viene così invitato a riformulare la frase facendo uso di strutture sintattiche più semplici, ad esempio evitando l'uso di una subordinata implicita espressa con il gerundio ("limitando di fatto la libertà e l'eguaglianza dei cittadini") incassata all'interno di una subordinata relativa ("che, limitando di fatto la libertà e l'eguaglianza dei cittadini, impediscono il pieno sviluppo della persona umana e...").

3. READ-IT e un esempio di semplificazione legislativa

Obiettivo di questo paragrafo è mostrare come le potenzialità di uno strumento di analisi della leggibilità come READ-IT possano essere utilizzate a supporto di un processo di *drafting* normativo che tenga conto delle indicazioni di semplicità, chiarezza ed efficacia comunicativa pensate per i testi della comunicazione istituzionale (cfr. § 1).

Abbiamo scelto come esempio la legge provinciale della Provincia autonoma di Bolzano n. 7 del 14 luglio 2015 che promuove la partecipazione e l'inclusione sociale delle persone con disabilità[36]. Coerentemente

[34] La versione a colori dell'immagine è disponibile alla pagina http://www.italianlp.it/wp-content/uploads/downloads/figure__readit/Figura-7.png.

[35] Per visualizzare nell'interfaccia web i valori di leggibilità relativi ad ogni livello è sufficiente passare con il cursore sulle colonne colorate corrispondenti.

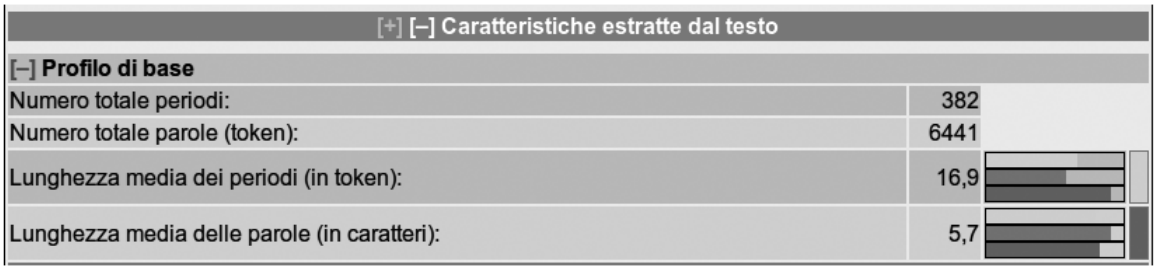
[36] http://www.regione.taa.it/bur/pdf/I-II/2015/37/BO/BO371501101570.PDF.

con gli obiettivi della legge, il 25 agosto 2015 la Giunta ne ha approvato anche una versione semplificata perché, come leggiamo nell'introduzione di questa versione, "tutte le persone devono capire cosa c'è scritto nella legge"[37]. L'analisi di questo testo ci è dunque parsa particolarmente significativa: l'attività di riscrittura semplificata rappresenta infatti un esempio virtuoso di semplificazione linguistica promosso in prima persona da una pubblica amministrazione.

La legge è stata riscritta non solo usando una *lingua facile*, ma adottando anche alcuni accorgimenti grafici volti a rendere più agevole la lettura del testo da parte di chi ha difficoltà. Il testo è dunque stato scritto con un carattere più grande, allineato a sinistra e sono state aggiunte delle immagini per spiegare anche in modo iconografico il contenuto della legge. È stato inoltre aggiunto un glossario finale nel quale si spiega il significato di parole di cui non era possibile trovare un sinonimo semplice. Dal punto di vista strettamente linguistico sono state messe in atto alcune strategie di semplificazione che hanno interessato più livelli di descrizione linguistica: lessicale, morfo-sintattica e sintattica.

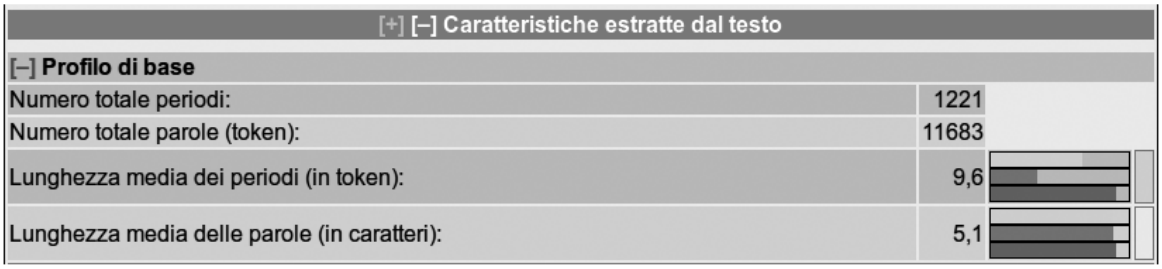
Allo scopo di illustrare le principali modifiche operate a livello linguistico sul testo originale, abbiamo analizzato la legge con READ-IT sia nella sua versione originale sia in quella riscritta. L'obiettivo era quello di dimostrare come lo strumento sia in grado di intercettare i luoghi di complessità linguistica del testo originale sui quali il gruppo di esperti che ha riscritto la legge è intervenuto allo scopo di renderla più comprensibile. Come mostreremo in quanto segue, READ-IT, mettendo in luce le differenze del profilo linguistico delle due versioni, non solo consente di fornire una valutazione assoluta del loro diverso livello di leggibilità ma riconosce anche le strutture lessicali, morfo-sintattiche e sintattiche che sono state modificate allo scopo di rendere più semplice la legge. Questo rende dunque READ-IT un valido ausilio alla semplificazione semi-automatica del testo.

[37] http://www.provincia.bz.it/politiche-sociali/download/LP_disabilita_in_lingua_facile.pdf.



[+] [–] Caratteristiche estratte dal testo	
[–] **Profilo di base**	
Numero totale periodi:	382
Numero totale parole (token):	6441
Lunghezza media dei periodi (in token):	16,9
Lunghezza media delle parole (in caratteri):	5,7

Figura 8. Caratteristiche linguistiche del *Profilo di base* della versione originale della legge provinciale n. 7 del 14 luglio 2015 della Provincia autonoma di Bolzano.



[+] [–] Caratteristiche estratte dal testo	
[–] **Profilo di base**	
Numero totale periodi:	1221
Numero totale parole (token):	11683
Lunghezza media dei periodi (in token):	9,6
Lunghezza media delle parole (in caratteri):	5,1

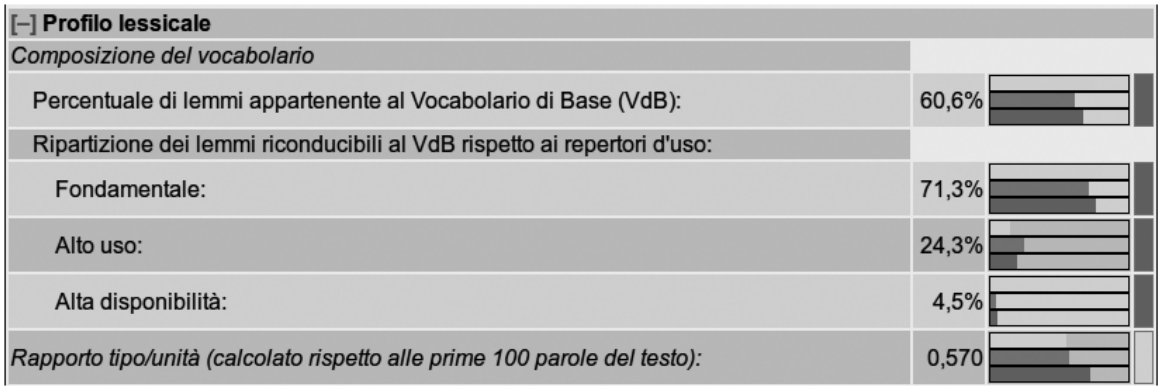
Figura 9. Caratteristiche linguistiche del *Profilo di base* della versione semplificata della legge provinciale n. 7 del 14 luglio 2015 della Provincia autonoma di Bolzano.

Come si può vedere nelle figure 8 e 9[38], una delle prime modifiche linguistiche operate sul testo semplificato riguarda la lunghezza della frase: il testo semplificato è stato riscritto con frasi mediamente più corte. Mentre infatti le frasi della versione originale della legge (figura 8) contengono una media di 16,9 token, quelle della versione semplificata (figura 9) sono più corte di circa 7 token. Per questo motivo il testo semplificato risulta molto più lungo dell'originale in termini di numero di frasi totali.

A livello del lessico sono state operate delle scelte volte a non snaturare la specificità di testo normativo che la legge mantiene anche nella sua versione semplificata. Come mostrano le figure 10 e 11[39], la versione semplificata

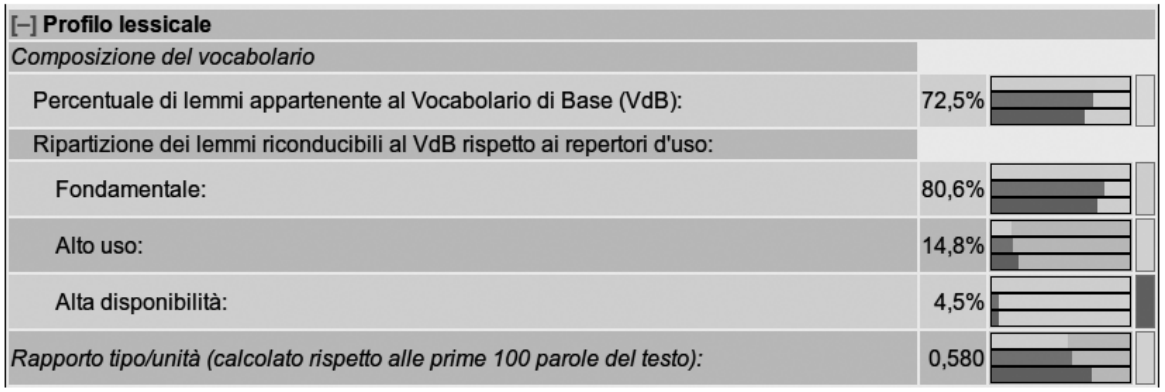
[38] Le versioni a colori delle Figure 8 e 9 sono disponibili rispettivamente alla pagina http://www.italianlp.it/wp-content/uploads/downloads/figure__readit/Figura-8.png e http://www.italianlp.it/wp-content/uploads/downloads/figure__readit/Figura-9.png.

[39] Le versioni a colori delle figure 10 e 11 sono disponibili rispettivamente alla pagina http://www.italianlp.it/wp-content/uploads/downloads/figure__readit/Figura-10.png e http://www.italianlp.it/wp-content/uploads/downloads/figure__readit/Figura-11.png.



[–] **Profilo lessicale**	
Composizione del vocabolario	
Percentuale di lemmi appartenente al Vocabolario di Base (VdB):	60,6%
Ripartizione dei lemmi riconducibili al VdB rispetto ai repertori d'uso:	
Fondamentale:	71,3%
Alto uso:	24,3%
Alta disponibilità:	4,5%
Rapporto tipo/unità (calcolato rispetto alle prime 100 parole del testo):	0,570

Figura 10. Caratteristiche linguistiche del *Profilo lessicale* della versione originale della legge provinciale n. 7 del 14 luglio 2015 della Provincia autonoma di Bolzano.



[–] **Profilo lessicale**	
Composizione del vocabolario	
Percentuale di lemmi appartenente al Vocabolario di Base (VdB):	72,5%
Ripartizione dei lemmi riconducibili al VdB rispetto ai repertori d'uso:	
Fondamentale:	80,6%
Alto uso:	14,8%
Alta disponibilità:	4,5%
Rapporto tipo/unità (calcolato rispetto alle prime 100 parole del testo):	0,580

Figura 11. Caratteristiche linguistiche del *Profilo lessicale* della versione semplificata della legge provinciale n. 7 del 14 luglio 2015 della Provincia autonoma di Bolzano.

della legge contiene una percentuale maggiore di lessico appartenente al *Vocabolario di Base* (pari al 72,5% di tutto il lessico della legge vs. il 60,6% della legge nella sua versione originale). Ad aumentare nel processo di semplificazione è stato soprattutto il lessico appartenente al repertorio dei lemmi di *Uso Fondamentale* e di *Alto Uso* (passati rispettivamente dal 71,3% della versione originale all'80,6% della versione semplificata e dal 24,3% al 14,8%). Non è variata invece la percentuale dei lemmi di *Alta Disponibilità* (in entrambe le versioni pari al 4,5% dei lemmi): come spiegato dai redattori stessi è stato necessario mantenere i termini più tecnici (e difficili) per non mutare la referenzialità del testo rispetto alla materia normativa trattata. Sono infatti questi i lemmi spiegati nel glossario aggiunto. Come si può notare, non è variato neanche il valore del *Rapporto tipo/unità*: è questo un

indice di ricchezza lessicale, ampiamente utilizzato in statistica lessicale, che oscilla tra 0 (testo con un vocabolario poco vario) e 1 (testo con vocabolario molto vario)[40]. La nostra legge, pertanto, con 0,570 nella versione originale e 0,580 nella versione semplificata risulta essere un testo non molto ricco lessicalmente. Il risultato non stupisce ed è in linea con studi linguistici che hanno dimostrato come i testi giuridici siano generalmente più lessicalmente poveri di altre varietà testuali (Nystedt 1999; Venturi 2012b). A tal proposito, ci interessa inoltre qui ricordare come la "Guida per la redazione degli atti amministrativi" (cfr. nota 10 del paragrafo introduttivo) raccomandi di "usare sempre il medesimo termine per esprimere uno stesso concetto; alternare termini diversi per indicare lo stesso concetto al fine di evitare le ripetizioni può generare confusione e ambiguità".

Ma è soprattutto a livello delle strutture sintattiche che i revisori della legge sono intervenuti di più. Alcuni dei tratti esaminati riflettono in parte quanto avevamo osservato a proposito della lunghezza della frase. Come si può notare dal confronto tra la figura 12 e 13[41], le frasi della versione originale, mediamente più lunghe di quelle contenute nella versione semplificata, sono le frasi che si caratterizzano anche per gli alberi sintattici più profondi, con una media di 5 relazioni di dipendenza consecutive per frase contro le 3 degli alberi sintattici delle frasi della versione semplificata. Come fatto notare in letteratura (Frazier 1985; Gibson 1998), è questo uno dei tratti sintattici maggiormente correlati con la complessità sintattica insieme alla lunghezza media delle relazioni di dipendenza sintattica, calcolata come la distanza tra la testa e il dipendente (in token). Rispetto a questo tratto, più una frase contiene relazioni di dipendenza lunghe più è difficile da comprendere, dal momento che è richiesto un maggiore impegno cognitivo (Miller 1956; Hudson 1995). La legge

[40] Misura ampiamente utilizzata in statistica lessicale, il Rapporto tipo/unità consiste nel rapporto tra il numero di parole tipo in un testo, il 'vocabolario' di un testo (V_c), e il numero delle occorrenza delle unità del vocabolario nel testo (C): $0 \leq | V/C_c | \leq 1$.

[41] Le versioni a colori delle figure 12 e 13 sono disponibili rispettivamente alla pagina http://www.italianlp.it/wp-content/uploads/downloads/figure__readit/Figura-12.png e http://www.italianlp.it/wp-content/uploads/downloads/figure__readit/Figura-13.png.

Struttura sintattica a dipendenze	
Articolazione interna del periodo:	
Numero medio di proposizioni per periodo:	0,974
Proposizioni principali vs subordinate (%)	
Principali:	82,6%
Subordinate:	17,4%
Articolazione interna della proposizione:	
Numero medio di parole per proposizione:	17,315
Numero medio di dipendenti per testa verbale:	1,890
"Misura" della profondità dell'albero sintattico:	
Media delle altezze massime:	4,949
Profondità media di strutture nominali complesse:	1,536
Profondità media di "catene" di subordinazione:	1,042
"Misura" della lunghezza delle relazioni di dipendenza (calcolata come distanza in parole tra testa e dipendente):	
Lunghezza media:	2,392
Media delle lunghezze massime:	7,301

Figura 12. Caratteristiche linguistiche del *Profilo sintattico* della versione originale della legge provinciale n. 7 del 14 luglio 2015 della Provincia autonoma di Bolzano.

nella versione originale contiene frasi con relazioni di dipendenza sintattica lunghe il doppio (con una lunghezza massima di 7 token) rispetto a quelle contenute nelle frasi della versione semplificata (pari a 3 token). I revisori della legge hanno inoltre agito su di un altro elemento tipicamente connesso con la complessità della lingua del diritto: la lunghezza media dei complementi preposizionali dipendenti in sequenza da un nome. Come fatto notare da Mortara Garavelli (2001), la propensione per l'uso di sostantivi per lo più astratti fa sì che siano "specialmente i nessi, i grappoli di astrazioni concatenate in 'complementi del nome' a marcare sintatticamente (e testualmente) gli enunciati". La conseguenza più significativa di tali "complicazioni strutturali" è che esse possono diventare "fonte di oscurità o di difficoltà interpretative". Nel processo di semplificazione è stata infatti ridotta la lunghezza media delle sequenze di complementi preposizionali (*Profondità media di strutture nominali complesse* nelle figure 12 e 13).

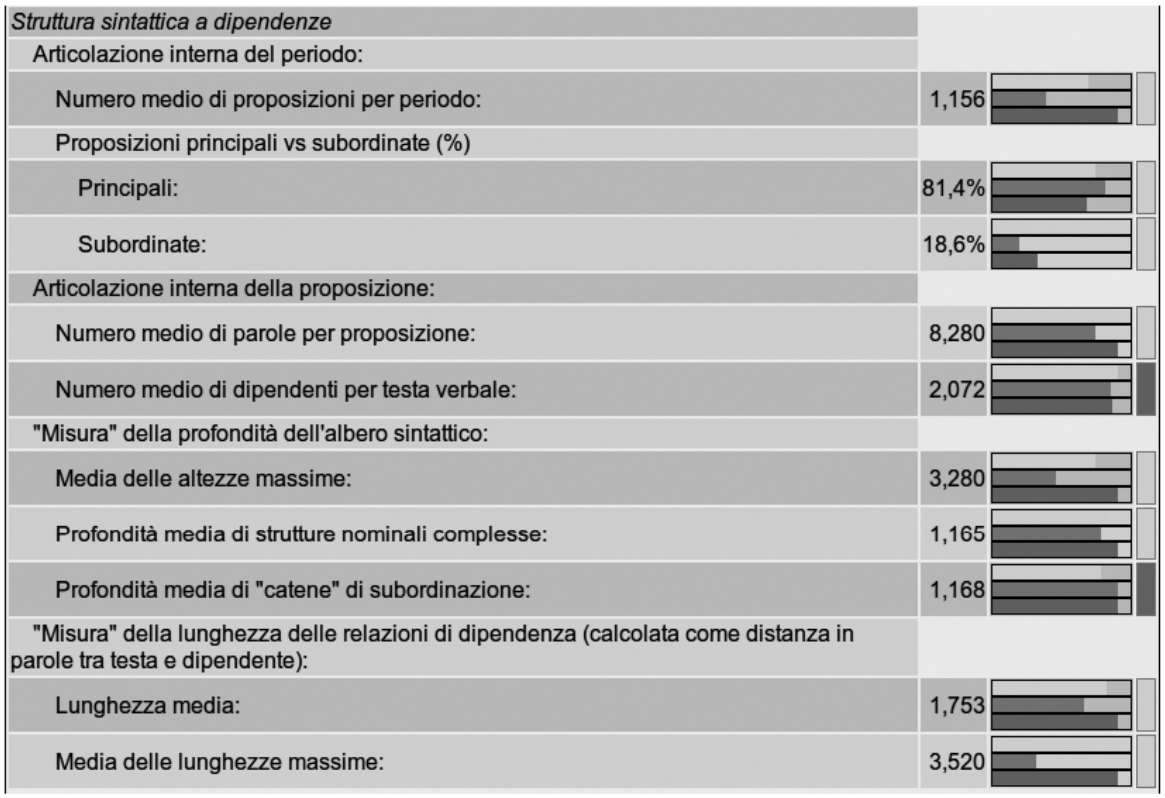


Figura 13. Caratteristiche linguistiche del *Profilo sintattico* della versione semplificata della legge provinciale n. 7 del 14 luglio 2015 della Provincia autonoma di Bolzano.

In linea con l'approccio impiegato da READ-IT per il calcolo della leggibilità, basato sui risultati del monitoraggio linguistico (vedi § 2.2), le modifiche delle strutture linguistiche descritte sopra si riflettono nei diversi punteggi di leggibilità. Come mostrano le figure 14 e 15[42], il processo di riscrittura della legge ha permesso di migliorare decisamente il suo livello di leggibilità *globale*. Mentre la versione originale aveva un valore di difficoltà di lettura del 99,6%, dopo la riscrittura il testo ha un valore di 29,1%. Come si può vedere, tutti i modelli di READ-IT sono interessati. La legge diventa più semplice sia rispetto al modello BASE, basato sulle caratteristiche del testo usate anche dagli indici tradizionali (cioè, lunghezza della frase e delle parole), sia rispetto al

[42] Le versioni a colori delle figure 14 e 15 sono disponibili rispettivamente alla pagina http://www.italianlp.it/wp-content/uploads/downloads/figure__readit/Figura-14.png e http://www.italianlp.it/wp-content/uploads/downloads/figure__readit/Figura-15.png.

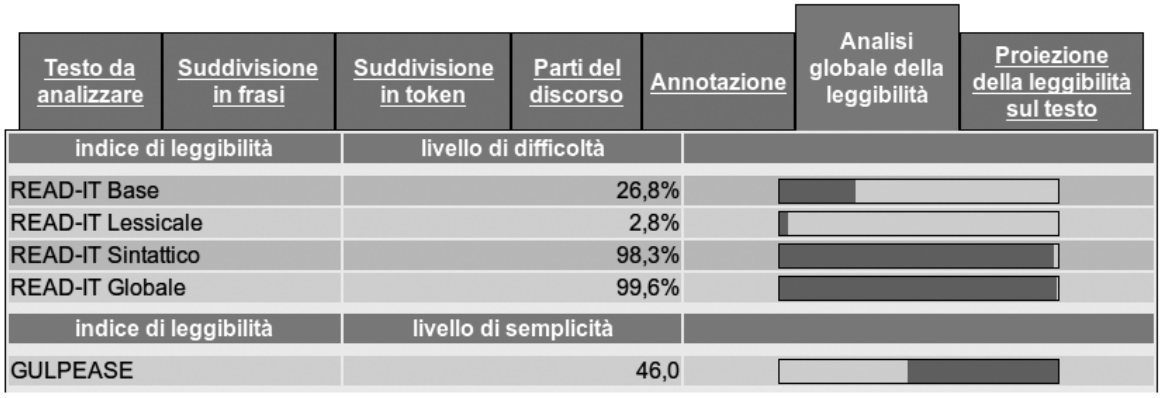


Figura 14. Risultato del calcolo globale della leggibilità della versione originale della legge provinciale n. 7 del 14 luglio 2015 della Provincia autonoma di Bolzano.

modello LESSICALE. Sebbene tuttavia rispetto alle caratteristiche linguistiche utilizzate da questi due modelli il testo originale non risultasse eccessivamente difficile, è il modello SINTATTICO che intercetta i maggiori cambiamenti fatti sul testo in fase di riscrittura. Come discusso prima, i redattori sono intervenuti infatti a livello delle strutture sintattiche del testo. L'indice di leggibilità basato sull'analisi di queste strutture è infatti quello che varia di più passando da un valore pari a 98,3% al 18,9%.

La portata applicativa delle indicazioni fornite da READ-IT diventa ancor più interessante nel momento in cui l'attenzione si sposta dall'analisi della leggibilità dell'intero testo a quella delle singole frasi (figure 16 e

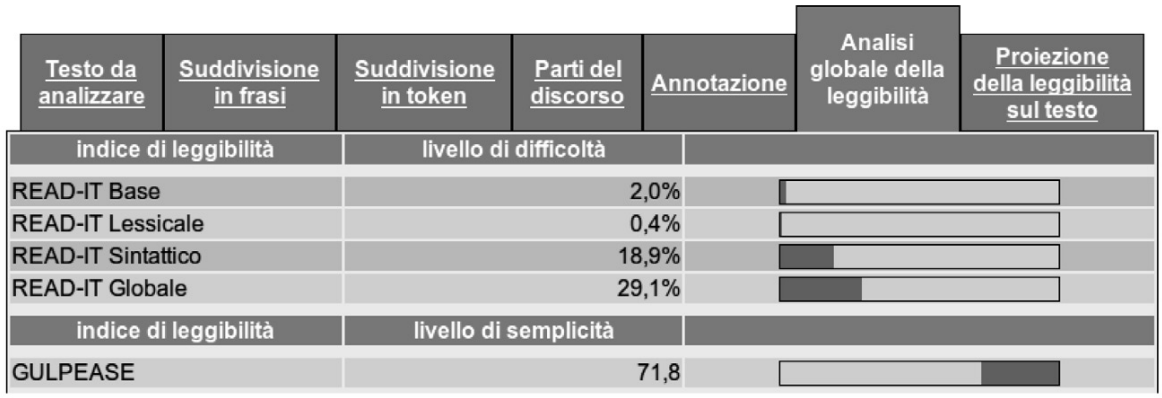


Figura 15. Risultato del calcolo globale della leggibilità della versione semplificata della legge provinciale n. 7 del 14 luglio 2015 della Provincia autonoma di Bolzano.

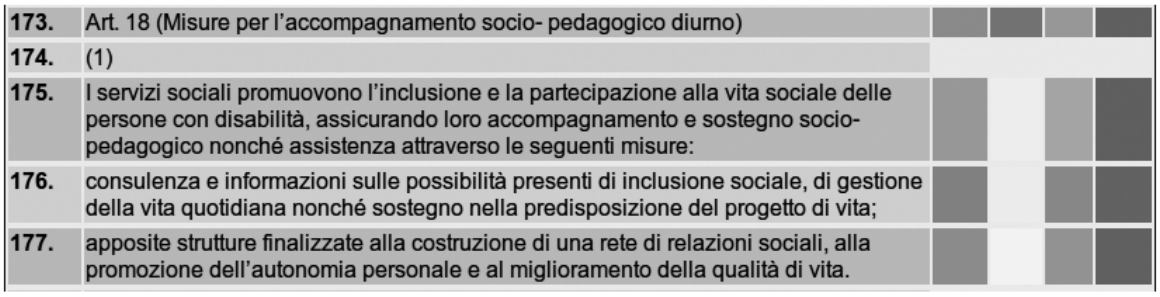


Figura 16. Leggibilità delle singole frasi della versione originale della legge provinciale n. 7 del 14 luglio 2015 della Provincia autonoma di Bolzano.

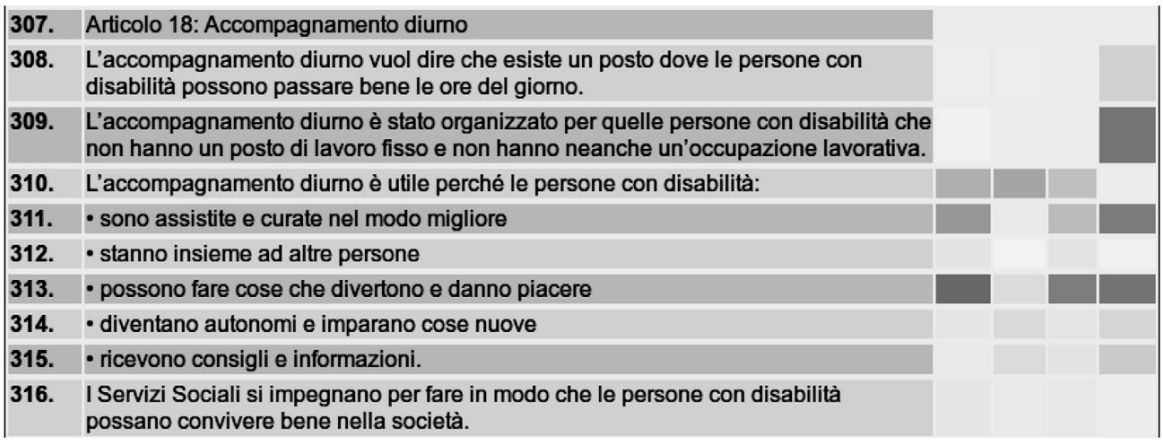


Figura 17. Leggibilità delle singole frasi della versione semplificata della legge provinciale n. 7 del 14 luglio 2015 della Provincia autonoma di Bolzano.

17[43]). Abbiamo scelto di mostrare l'analisi delle frasi dell'articolo 18 della legge che descrive le Misure per l'accompagnamento socio-pedagogico diurno perché rappresentano un esempio significativo di come il processo di riscrittura semplificata abbia interessato di più le strutture sintattiche del testo mantenendo invece alcune scelte lessicali che non potevano essere modificate.

È stato scelto innanzitutto di accorciare le due frasi cuore dell'articolo che spiegano in cosa consistano le misure di accompagnamento (frasi

[43] Le versioni a colori delle figure 16 e 17 sono disponibili rispettivamente alla pagina http://www.italianlp.it/wp-content/uploads/downloads/figure__readit/Figura-16.png e http://www.italianlp.it/wp-content/uploads/downloads/figure__readit/Figura-17.png.

176 e 177 della versione originale). Inoltre, sono stati sciolti i numerosi costrutti nominali presenti nella formulazione originale (es. consulenza, gestione, sostegno, predisposizione). È stata invece adottata una struttura ad elenco – più chiara anche nella resa grafica – introdotta da una frase sintatticamente molto semplice (310): soggetto, predicato nominale e subordinata causale. Le singole misure sono poi elencate nelle successive cinque frasi autonome dove ciascuna informazione è veicolata da costruzioni grammaticali semplici in cui l'azione è sempre resa in maniera esplicita dal verbo.

L'intervento di semplificazione lessicale è stato invece minore dal momento che alcuni termini come "accompagnamento diurno" non potevano essere eliminati. Sono stati presi altri tipi di accorgimenti: è stato aggiunto il termine nel glossario conclusivo, è stato spiegato nelle prime due frasi dell'articolo (frasi 308 e 309 della versione semplificata) cosa si intende per "accompagnamento diurno" e a quali persone è rivolto (due frasi che hanno un livello di leggibilità sintattica di 43,4% e 42,9%).

4. Conclusioni e sviluppi futuri

Una comunicazione istituzionale di qualità è una comunicazione chiara, semplice ed efficace ed è senza dubbio uno dei principali segnali di efficienza di una pubblica amministrazione. La possibilità di valutare la leggibilità dei testi istituzionali rappresenta un passo fondamentale verso una comunicazione pubblica amministrazione-cittadini più accessibile. Una metodologia come quella qui presentata basata su strumenti di Trattamento Automatico del Linguaggio rappresenta un valido ausilio in questa direzione. Tale metodologia, già sperimentata con successo su diverse tipologie di testi, si è rivelata affidabile anche nel caso dei testi rappresentativi della alla comunicazione istituzioni-cittadini qui analizzati.

Lo studio ha permesso di dimostrare come tecnologie linguistiche allo stato dell'arte per la lingua italiana siano oggi mature non solo per calcolare il grado di leggibilità di testi istituzionali ma anche come guida per la loro stesura semplificata. Sebbene infatti la metodologia di monitorag-

gio linguistico e analisi della leggibilità sia stata originariamente messa a punto per il trattamento di testi rappresentativi della lingua comune, gli esperimenti condotti hanno dimostrato come essa riesca tuttavia ad intercettare le difficoltà della *lingua del diritto*, mettendone in luce gli specifici luoghi di complessità. Nell'esempio di semplificazione della legge regionale della Provincia Autonoma di Bolzano qui discusso abbiamo mostrato come READ-IT permetta di riconoscere le strutture lessicali, morfo-sintattiche e sintattiche che sono state modificate dai redattori allo scopo di rendere la legge più semplice. Questo grazie all'innovativa possibilità di READ-IT, che permette di valutare la leggibilità non solo di un intero documento ma anche di ogni singola frase che lo compone. L'indicazione dei principali luoghi di complessità della frase è infatti il primo passo verso la definizione di una metodologia di semplificazione semi-automatica del testo. Processo di semplificazione tanto più centrale in una reale società inclusiva dove il rapporto istituzioni-cittadini dovrebbe essere al centro della vita democratica di uno Stato.

Oltre a fornire un valido supporto alla semplificazione assistita del testo, tra le possibili direzioni di ricerca offerte dalla metodologia di analisi della leggibilità e semplificazione del testo qui presentata vi è la sua adozione in un contesto di formazione dei funzionari delle pubbliche amministrazioni. Da un'inchiesta finalizzata a fotografare come i dipendenti pubblici recepiscano le sollecitazioni alla semplificazione del linguaggio amministrativo nella pratica di scrittura quotidiana (Viale 2008) è emerso infatti che l'"abbandono" del burocratese spesso non è dovuto ad una scarsa sensibilità del dipendente pubblico ai temi della semplificazione linguistica, quanto ad un'organizzazione del lavoro che porta a riciclare modelli di scrittura preconfezionati nella convinzione, errata, che esista un insuperabile divario tra la lingua comune e quella richiesta nei testi amministrativi. Queste riflessioni sono importanti perché testimoniano quanto la formazione linguistica del dipendente pubblico svolga un ruolo di primo piano. Questa formazione include varie competenze, tra cui la capacità di saper padroneggiare i diversi registri linguistici in funzione del destinatario e saper riconoscere i contesti d'uso formale di un termine o di una costruzione grammaticale. Nella prospettiva di diffondere nella

pratica di chi scrive modelli comunicativi ispirati alla chiarezza e alla semplificazione linguistica, uno strumento come READ-IT potrebbe essere proposto nell'ambito di corsi di formazione rivolti al personale amministrativo, affiancando i più tradizionali editoriali dedicati alla semplificazione del linguaggio istituzionale.

Bibliografia

Aluisio S., L. Specia, C. Gasperin and C. Scarton, *Readability Assessment for Text Simplification. Proceedings of the NAACL HLT 2010 Fifth Workshop on Innovative Use of NLP for Building Educational Applications*, Association for Computational Linguistics, 2010, pp. 1-9.

Attardi G., F. Dell'Orletta, M. Simi and J. Turian, *Accurate Dependency Parsing with a Stacked Multilayer Perceptron*, in *Proceedings of Evalita '09 (Evaluation of NLP and Speech Tools for Italian)*, Reggio Emilia, 2009, disponibile alla pagina: http://www.evalita.it/sites/evalita.fbk.eu/files/proceedings2009/Parsing/Dependency/DEP_PARS_UNIPI_UNI_MONTREAL.pdf

Brown G.D. and F.L. Watson, *First in, first out: Word learning age and spoken word frequency as predictors of word familiarity and word naming latency*, in *Memory & Cognition*, 15-3/1987, pp. 208-216.

Calvino I., *Per ora sommersi dall'antilingua*, in *Il Giorno*, 3 febbraio 1965 (ora in *Una pietra sopra*, Torino 1980, pp. 122-126).

Cortelazzo M.A. e F. Pellegrino, *Guida alla scrittura istituzionale*, Roma-Bari 2003.

Curtotti M. and E. McCreath, *A Right to Access Implies A Right to Know: An Open Online Platform for Research on the Readability of Law*, in *Journal of Open Access to Law*, 1-1/2013, pp. 1-56.

Cutts M., *The Plain English Guide*, Oxford 1995.

De Mauro T., *Grande dizionario italiano dell'uso (GRADIT)*, Torino 2000.

De Mauro T., *Introduzione. Il linguaggio della Costituzione*, in *Costituzione della Repubblica Italiana (1947)*, Torino 2006, pp. vii-xxxii.

Dell'Orletta F., *Ensemble system for Part-of-Speech tagging*, in *Proceedings of Evalita '09 (Evaluation of NLP and Speech Tools for Italian)*, Reggio Emilia 2009, disponibile alla pagina http://www.evalita.it/sites/evalita.fbk.eu/files/proceedings2009/PoSTagging/POS_ILC.pdf.

Dell'Orletta F. e S. Montemagni, *Tecnologie linguistico-computazionali per la valutazione delle competenze linguistiche in ambito scolastico*, in *Atti del XLIV Congresso Internazionale di Studi della Società di Linguistica Italiana (SLI 2010)*, Viterbo, 27-29 settembre 2010.

Dell'Orletta F., S. Montemagni, E.M. Vecchi e G. Venturi, *Tecnologie linguistico-computazionali per il monitoraggio della competenza linguistica italiana degli alunni stranieri nella scuola primaria e secondaria*, in G.C. Bruno, I. Caruso, M. Sanna e I. Vellecco (a cura di), *Percorsi migranti: uomini, diritto, lavoro, linguaggi*, Milano 2011a, pp. 319-336.

Dell'Orletta F., S. Montemagni and G. Venturi G., *READ-IT: assessing readability of Italian texts with a view to text simplification*, in *Proceedings of the Second Workshop on Speech and Language Processing for Assistive Technologies (SLPAT '11)*, Edimburgo, UK, 30 luglio 2011b, pp. 73-83.

Dell'Orletta F., S. Marchi, S. Montemagni, B. Plank and G. Venturi, *The SPLeT-2012 Shared Task on Dependency Parsing of Legal Texts*, in *Proceedings of the LREC 2012 4th Workshop on Semantic Processing of Legal Texts (SPLeT 2012)*, Istanbul, Turkey, 27 May 2012.

Dell'Orletta F., S. Marchi, S. Montemagni, G. Venturi, T. Agnoloni and E. Francesconi, *Domain Adaptation for Dependency Parsing at Evalita 2011*, in B. Magnini, F. Cutugno, M. Falcone and E. Pianta (ed. by), *Evaluation of Natural Language and Speech Tool for Italian*, LNCS-LNAI, vol. 7689, Berlin Heidelberg 2013a, pp. 58-69.

Dell'Orletta F., S. Montemagni and G. Venturi, *Linguistic Profiling of Texts Across Textual Genre and Readability Level. An Exploratory Study on Italian Fictional Prose*, in *Proceedings of the Recent Advances in Natural Language Processing Conference (RANLP-2013)*, Hissar, Bulgaria, 7-11 September 2013b, pp. 189-197.

Dell'Orletta F., S. Montemagni and G. Venturi, *Assessing Document and Sentence Readability in Less Resourced Languages and across Textual Genres*, in T. François and D. Bernhard (ed. by), *International Journal of Applied Linguistics (ITL). Special Issue on Readability and Text Simplification*, 2014a, pp. 163-193.

Dell'Orletta F., M. Wieling, G. Venturi, A. Cimino and S. Montemagni, *Assessing the Readability of Sentences: Which Corpora and Features?*, in *Proceedings of the Ninth Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2014)*, Baltimore, Maryland, 26 June 2014b, pp. 163-173.

Ferreira F., *The misinterpretation of noncanonical sentences*, in *Cognitive Psychology*, 47/2003, pp. 164-203.

Fioritto A. (a cura di), *Manuale di Stile. Strumenti per semplificare il linguaggio delle amministrazioni pubbliche*, Bologna 1997.

Fortis D., *Il linguaggio amministrativo italiano*, in *Revista de Liengua i dret*, 43/2005, pp. 47-116.

Fortis D., *Il Plain Language: quando le istituzioni si fanno capire*, I quaderni del Mestiere di Scrivere, 2003.

Franceschini F. e S. Gigli (a cura di), *Manuale di scrittura amministrativa*, Roma 2003.

Frazier L., *Syntactic complexity*, in D.R. Dowty, L. Karttunen and A.M. Zwicky (ed. by), *Natural Language Parsing*, Cambridge 1985.

Gibson E., *Linguistic complexity: Locality of syntactic dependencies*, in *Cognition*, 68-1/1998, pp. 1-76.

Gildea D., *Corpus variation and parser performance*, in *Proceedings of Empirical Methods in Natural Language Processing (EMNLP 2001)*, Pittsburgh, PA 2001, pp. 167-202.

Hudson R., *Measuring syntactic difficulty*, unpublished paper, 1995, consultabile alla pagina: http://dickhudson.com/wp-content/uploads/2013/07/Difficulty.pdf.

Katz D.M. and M.J. Bommarito, *Measuring the Complexity of the Law: The United States Code (August 1, 2013)*, 2014, consultabile alla pagina: http://ssrn.com/abstract=2307352

Kincaid J.P., R.P. Fishburne, R.R. Rogers and B.S. Chissom, *Derivation of new readability formulas for Navy enlisted personnel*, Research Branch Report, Millington, TN 1975, pp. 8-75.

Lucisano P. e M.E. Piemontese (1988), *Gulpease. Una formula per la predizione della difficoltà dei testi in lingua italiana*, in *Scuola e Città*, 3/1988, pp. 57-68.

Marinelli R., L. Biagini, R. Bindi, S. Goggi, M. Monachini, P. Orsolini, E. Picchi, S. Rossi, N. Calzolari and A. Zampolli, *The Italian PAROLE corpus: an overview*, in A. Zampolli et al. (ed. by), *Computational Linguistics in Pisa*, XVI-XVII(1), Pisa-Roma 2003, pp. 401-421.

Mercatali P. e F. Romano (2013), *I documenti dello stato digitale. Regole e tecniche per la semplificazione*, collana d'informatica giuridica, 2/2013.

Miller G., *The magical number seven, plus or minus two: some limits on our capacity for processing information*, in *Psycological Review*, 63/1956, 81-97.

Montemagni S., *Tecnologie linguistico-computazionali e monitoraggio della lingua italiana*, in *Studi Italiani di Linguistica Teorica e Applicata (SILTA)*, XLII-1/2013, pp. 145-172

Mortara Garavelli B., *Le parole e la giustizia. Divagazioni grammaticali e retoriche su testi giuridici italiani*, Torino 2001.

Nystedt J., *Ricchezza (o povertà?) lessicale nei documenti italiani della CEE*, in G. Alfieri, A. Cassola (a cura di), *La "Lingua d'Italia". Usi pubblici e istituzionali*, Atti del XXlX Congresso Internazionale di Studi della SLI (Malta, 3-5 novembre 1998), Roma 1999, pp. 471-491.

Piemontese M.E. e M.T. Tiraboschi, *Leggibilità e comprensibilità dei testi della pubblica amministrazione. Strumenti e metodologie di ricerca al servizio del diritto a capire testi di rilievo pubblico*, in E. Zuanelli (a cura di), *Il diritto all'informazione in Italia*, Roma 1990, pp. 225-246.

Piemontese M.E., *Capire e farsi capire. Teorie e tecniche della scrittura controllata*, Napoli 1996.

Piemontese M.E., *Il linguaggio della pubblica amministrazione nell'Italia d'oggi. Aspetti problematici della semplificazione linguistica*, in G. Alfieri e A. Cassola (a cura di), *La "Lingua d'Italia". Usi pubblici e istituzionali*, Atti del XXlX Congresso Internazionale di Studi della SLI (Malta, 3-5 novembre 1998), Roma 1999, pp. 269-292.

Piemontese M.E., *Leggibilità e comprensibilità delle leggi italiane. Alcune osservazioni quantitative e qualitative*, in D. Veronesi (a cura di), *Linguistica giuridica italiana e tedesca: obiettivi, approcci, risultati*, Atti del Convegno di studi (Bolzano, 1-3 ottobre 1998), Padova 2000, pp. 103-117.

Piemontese M.E. (2001), *Leggibilità e comprensibilità dei testi delle pubbliche amministrazioni: problemi risolti e problemi da risolvere*, in S. Covino (a cura di), *La scrittura professionale. Ricerca, prassi, insegnamento*, Atti del Convegno di studi (Perugia, Università per stranieri, 23-25 ottobre 2000), Firenze 2001, pp. 119-130.

Raso T., *La scrittura burocratica. La lingua e l'organizzazione del testo*, Roma 2005.

Segui J., J. Mehler, U. Frauenfelder and J. Morton, *The word frequency effect and lexical access*, in *Neuropsychologia*, 20/1982, pp. 615-627.

Serianni L., *Un treno di sintomi. I medici e le parole. Percorsi linguistici nel passato e nel presente*, Milano 2005.

Venturi G., *Design and Development of TEMIS: a Syntactically and Semantically Annotated Corpus of Italian Legislative Texts*, in *Proceedings of the LREC 2012 4th Workshop on "Semantic Processing of Legal Texts"*, Istanbul, Turkey, 27 May 2012a, pp. 1-12.

Venturi G., *Investigating legal language peculiarities across different types of Italian legal texts: an NLP-based approach*, in *Bridging the gap(s) between Language and the Law. Proceedings of the 3rd European Conference of the International Association of Forensic Linguistics*, Porto, Portogallo, 15-18 ottobre 2012b, pp. 138-156.

Viale M., *Studi e ricerche sul linguaggio amministrativo*, Padova 2008.

Zuanelli E. (a cura di), *Il diritto all'informazione in Italia*, Roma 1990.