



UNIVERSITÀ DI PISA

Corso di Laurea in Informatica Umanistica

RELAZIONE

Definizione di un metodo e creazione di un corpus
per la valutazione di un sistema di semplificazione
automatica del testo

Candidato: *Giulia Cantoni*

Relatore: *Felice Dell'Orletta*

Dominique Brunato

Correlatore: *Alessandro Lenci*

Anno Accademico 2016-2017

Indice

1	Introduzione	1
2	Stato dell'arte dei sistemi per valutare la text simplification	5
2.1	Metodi e parametri	5
2.2	Approccio umano	7
2.2.1	Applicazione del Crowdsourcing nella valutazione del testo semplificato	8
2.3	L'approccio ibrido: un esempio per la lingua spagnola	9
2.3.1	Fase preliminare: costruzione ed annotazione di un corpus parallelo monolingue	10
2.3.2	Fase di valutazione della Text Simplification: misure qualitative ed automatiche	11

3	Il corpus	13
3.1	Il Corpus del progetto	14
3.1.1	INVALSI e altri testi	14
3.2	Descrizione del Corpus	15
3.2.1	Analisi qualitativa del Corpus	16
3.2.2	Le domande e l'analisi del testo	21
3.2.3	Analisi della complessità linguistica	24
3.2.3.1	READ-IT	25
3.2.3.2	Risultati	27
4	Analisi del testo semplificato	34
4.1	Analisi dei risultati di READ-IT	34
4.2	Analisi dei risultati svolte manualmente	45
4.2.1	Analisi della grammaticalità	45
4.2.2	Analisi dell'informatività	48
5	Conclusioni	55
	Bibliografia	58
	Siti Web	59

1. INTRODUZIONE

Il focus di questo lavoro è la definizione di un metodo per la valutazione di un sistema di semplificazione automatica del testo. Questo metodo si propone di valutare, in particolare, una regola di semplificazione che consiste nella cancellazione (delete) di elementi (singole parole o frasi) in una frase. A questo scopo è stato costruito un corpus formato da testi appartenenti a diverse classi scolastiche delle scuole elementari, medie e superiori; a questi testi sono associate delle domande di comprensione del testo, che ci serviranno per la valutazione della Text simplification. Il corpus che è stato costruito a questo scopo e le regole che sono state applicate per la valutazione della semplificazione sono generali, possono cioè essere utilizzate per l'analisi e per la valutazione di qualsiasi regola di semplificazione.

La semplificazione automatica del testo fa parte del NLP (Natural Language Processing) e tenta di risolvere automaticamente il problema della complessità del lessico e della sintassi di una frase; le regole sulla semplificazione automatica del testo sono nate sia per rendere i contenuti elettronici testuali accessibili a tutti, sia per agevolare altri compiti del NLP. Questo perché con l'avvento di Internet la persona media ha accesso a molte più informazioni rispetto al passato e poter far parte della società attraverso l'accesso alle informazioni è un diritto essenziale di tutti; tuttavia numerose persone hanno difficoltà a leggere rapporti governativi, leggi, notizie, articoli e altri materiali scritti che consentono l'inserimento nella comunità, per questo alcuni portali internet hanno creato varianti semplificate dei loro contenuti, come nel caso della Simple English Wikipedia (https://simple.wikipedia.org/wiki/Main_Page).

Un modo per affrontare questo problema è l'adeguamento manuale dei contenuti testuali esistenti per le persone con esigenze particolari; semplificare manualmente un testo però è costoso e richiede competenze specifiche, e non è conveniente soprattutto quando si

tratta di articoli di cronaca, che vengono generati continuamente ed aggiornati. Ecco perché si è deciso di provare ad automatizzare il processo di semplificazione del testo. Tuttavia questo tipo di semplificazione è complessa perché comprende un gran numero di operazioni applicate a diversi livelli linguistici.

L' Automatic Text Simplification ha quindi lo scopo di trasformare un testo complesso in una sua variante semplificata, tenendo conto delle esigenze specifiche di una particolare classe di utenti di destinazione.

A partire dall'analisi di corpora paralleli (originali/semplificati) costruiti in varie lingue, sono state individuate alcune regole che chi semplifica un testo applica, in base all'utente di riferimento, per rendere un testo difficile un testo semplice. L'individuazione di queste regole si propone come un passo preliminare per riuscire a riprodurre queste regole in maniera automatica.

Alcune delle regole individuate per l'italiano ad esempio sono:

- **Split:** L'operazione di *split* ovvero di *divisione*, consiste nel segmentare una frase in più parti, in modo da avere frasi più brevi per esprimere gli stessi concetti, evitando al lettore un notevole sforzo di comprensione dovuto all'uso di frasi molto lunghe;
- **Merge:** L'operazione di *merge*, ovvero di *fusione*, è l'opposto dello split; consiste dunque nell'unione di due o più frasi in un'unica frase semplificata. Non è una regola molto utilizzata in quanto formare frasi molto lunghe potrebbe creare problemi di comprensione;
- **Reordering o Riordinamento:** L'operazione di *reordering* consiste nel cambiare l'ordine di alcune parti della frase originale, stando ovviamente attenti ad intervenire anche a livello di lessico e di sintassi (cambiare solo l'ordine potrebbe produrre frasi agrammaticali). È una regola utilizzata più frequentemente delle precedenti;

- **Insert o Inserimento:** L'operazione di *insert*, ci permette di aggiungere informazioni utili a comprendere meglio una frase. È un'operazione complessa per essere utilizzata da un sistema automatico;
- **Delete o Cancellazione:** L'operazione di *delete*, consiste nell'eliminazione di parole ridondanti o poco utili alla comprensione dell'evento descritto dalla frase. Come per l'insert, il delete risulta un'operazione difficile da eseguire automaticamente, in quanto l'identificazione delle espressioni da eliminare non è scontata. Può infatti coinvolgere aggettivi, verbi, avverbi, soggetti e interi sintagmi;
- **Transformation o Trasformazione:** è la regola più applicata, sono stati individuati 7 diversi generi di trasformazione: sostituzione lessicale a livello di parola, sostituzione lessicale a livello di espressione o sintagma, sostituzione di un pronome con il suo antecedente, nome in verbo, verbo in nome (nominalizzazione), diatesi, tratti verbali.

Queste regole possono essere applicate separatamente, tutte insieme o in modo combinato a seconda della strategia di semplificazione, del tipo di testo e del target di riferimento.

In questa tesi ci soffermeremo, come detto in precedenza, su una di queste regole in particolare: il Delete infatti, studi condotti su corpora paralleli in italiano hanno evidenziato che è una delle regole maggiormente applicate. Lo scopo del lavoro è soprattutto valutarne l'efficacia rispetto a diversi parametri, a partire da un corpus contenente domande di comprensione del testo.

La tesi è organizzata in questo modo: nel capitolo successivo parleremo dello stato dell'arte dei sistemi per valutare la Text simplification, vedremo quindi i metodi e i parametri utilizzati per la valutazione della semplificazione.

Nel terzo capitolo andremo ad analizzare il corpus che abbiamo costruito prendendo in analisi i testi da cui è formato, e le domande associate ai testi; faremo un'analisi preventiva delle frasi e dei testi non ancora semplificati rispetto al livello di complessità iniziale, utilizzando uno strumento chiamato READ-IT, per poterla poi confrontare con quella delle frasi semplificate.

Per quanto riguarda l'analisi del testo semplificato, la tesi verte sul valutare la semplificazione del testo e, come detto in precedenza, in modo particolare sulla semplificazione avvenuta applicando la regola: delete.

Nel capitolo cinque descriveremo il metodo di valutazione che combina approcci automatici e manuali. Valuteremo tre parametri: 1) la complessità delle frasi dopo la semplificazione, utilizzando READ-IT; 2) il grado di informatività delle frasi dopo la semplificazione, utilizzando le domande di cui abbiamo parlato nel capitolo precedente; 3) la loro grammaticalità. Confronteremo i dati di complessità analizzati prima della semplificazione con quelli ottenuti dopo la semplificazione.

2. STATO DELL'ARTE DEI SISTEMI PER VALUTARE LA TEXT SIMPLIFICATION

Una delle principali sfide in ambito di semplificazione automatica del testo è proprio la sua valutazione, che spesso si basa su misure di leggibilità. Tradizionalmente queste misure vengono calcolate automaticamente e tengono conto delle caratteristiche superficiali del testo, come il numero delle parole e il numero delle frasi.

Questo capitolo vuole porsi come un'introduzione alla valutazione del testo semplificato e discuterà sia i principali metodi di valutazione del testo semplificato, sia i parametri su cui si basa la valutazione stessa, questo perché la valutazione di un testo semplificato infatti non è una semplice valutazione basata sulla correttezza ortografica, sul lessico o sulla conoscenza da parte di chi scrive dell'argomento trattato.

Tra i metodi discussi si dedicherà un approfondimento alla valutazione della text simplification tramite Crowdsourcing e ad un esempio di valutazione di testi semplificati condotto sulla lingua spagnola.

2.1 Metodi e parametri

Per quanto riguarda i metodi di valutazione del testo semplificato si possono distinguere due grandi categorie: metodi automatici e metodi qualitativi.

Per quanto riguarda i metodi automatici parliamo ad esempio di metodi che si basano sull'applicazione di indici di leggibilità. I metodi automatici solitamente sono applicati

solo a livello del testo e non a livello delle frasi, contrariamente ai metodi qualitativi, che invece sono stati progettati proprio per essere applicati a livello delle frasi.

I metodi qualitativi sono quelli che si basano sul giudizio umano che viene ottenuto tramite metodi di Crowdsourcing o nella maggior parte dei casi rivolgendosi a professionisti. Il Crowdsourcing può essere utilizzato per raccogliere input da persone non esperte, quindi non interrogando professionisti ma chiunque voglia rispondere, al fine di fornire rapidamente una grande quantità di dati utili a valutare la semplicità del testo. I metodi qualitativi forniscono un modo per andare oltre le misure automatiche, che possono trascurare importanti modifiche al contenuto che di fatto aggiungono difficoltà di lettura del testo o eliminano informazioni rilevanti nel testo semplificato.

Per quanto riguarda i parametri su cui si basa la valutazione del testo semplificato parliamo di: grammaticalità, informatività e semplicità.

La grammaticalità, come dice la parola stessa, ci permette di valutare se il testo rispetta le regole grammaticali per quanto riguarda singole frasi o il testo nel suo insieme. Si valutano ad esempio l'uso della punteggiatura, i rapporti di reggenza, l'accordo tra articolo e nome, tra nome e aggettivo, tra soggetto e predicato ecc.

Per quanto riguarda la valutazione della semplicità di un testo vengono utilizzati spesso i metodi automatici, in quanto sono molto rapidi e nella maggior parte dei casi riescono a dare risultati positivi; per l'applicazione di questi metodi vengono ad esempio considerati il numero di parole del testo di partenza e il numero di parole del testo semplificato o il numero delle frasi, per valutare se un testo dopo la semplificazione è o non è più semplice di quello di partenza. Secondo questi criteri automatici si può dire che un testo è più semplice quando è più corto, ma non è sempre così: ad esempio, dal momento che l'esito della semplificazione dipende anche dal lettore di riferimento, un testo può essere corto, ma ricco di parole difficili o strutture sintattiche complesse per un utente con difficoltà cognitive. La semplicità di un testo non riguarda quindi solo la sua lunghezza, ma anche il suo lessico e la sua struttura sintattica.

L'informatività di un testo semplificato è un parametro su cui si basa per valutare se il testo dopo la semplificazione ha conservato il significato originale o se invece ha perso informazioni necessarie alla comprensione.

Per la valutazione di un testo è necessario utilizzare tutti e tre i parametri sopra elencati in modo da riuscire ad ottenere una valutazione esaustiva del cambiamento avvenuto al testo semplificato. Non basta quindi utilizzare un solo parametro per valutare la corretta semplificazione perché un parametro non esclude l'altro: un testo semplificato può rispettare tutte le regole grammaticali, ma può aver perso la sua informatività; può rispettare le regole grammaticali ed aver mantenuto la sua informatività, essere corto, ma contenere un lessico troppo difficile per essere letto da utenti con difficoltà cognitive e così via.

2.2 Approccio umano

L'approccio umano alla valutazione di un testo avviene, solitamente, utilizzando un gruppo di esperti; questo tipo di valutazione è piuttosto costosa e gli esperti potrebbero anche non essere disponibili quando servono, quindi può risultare un approccio per il quale serve una grande quantità di tempo. Per superare questi problemi si è iniziato in alcuni casi ad utilizzare un tipo di approccio umano differente: il Crowdsourcing.

Il Crowdsourcing è lo sviluppo collettivo di un progetto da parte di numerose persone esterne al promotore del progetto, che decidono volontariamente di partecipare rispondendo ad un invito a collaborare. Questo modello di realizzazione dei progetti è generalmente reso possibile da Internet e può riguardare una vasta varietà di progetti come ad esempio Wikipedia che è scritta dai propri lettori. Inizialmente il Crowdsourcing si basava sul lavoro di volontari ed appassionati che dedicavano il loro tempo libero a creare contenuti o a risolvere problemi, la prima a trarne beneficio è stata la community open source. Oggi grazie al Crowdsourcing i lavoratori freelance hanno la possibilità di offrire i propri servizi su un mercato globale offrendo dati utili anche a

scopi di ricerca ad esempio quella linguistica, sia teorica, sia pratica, dando modo di raccogliere quantità di dati importanti e fare nuove scoperte.

2.2.1 Applicazione del Crowdsourcing nella valutazione del testo semplificato

Walter S. Lasecki, Luz Rello e Jeffrey P. Bigham, in uno studio del 2015 hanno usato il Crowdsourcing per valutare l'output di un sistema di semplificazione automatica del testo. Il loro lavoro utilizza quindi il Crowdsourcing come mezzo rapido per valutare la semplicità testuale.

Di solito per la valutazione della semplicità dei testi viene utilizzato il giudizio di esperti, ma questo tipo di valutazione è piuttosto costosa in termini di tempo, risorse e competenze specialistiche; la presenza immediata di utenti è particolarmente importante per i sistemi di semplificazione automatica che hanno bisogno di un feedback immediato sulla semplicità della loro produzione. Il Crowdsourcing è una valida alternativa alla valutazione da parte di esperti, in quanto ci dà una valutazione rapida e meno costosa.

Per raccogliere i feedback dati dai partecipanti sulla semplicità dei testi è stata creata una semplice interfaccia che chiede ai lavoratori reclutati tramite la piattaforma Amazon Mechanical Turk di valutare il livello di semplicità di 10 frasi su una scala da 7 (semplice) a 1 (difficile). Le frasi che sono state date agli annotatori sono state semplificate sulla base di caratteristiche di complessità linguistica validate sperimentalmente da studi psico-linguistici che hanno evidenziato come la lunghezza delle parole e delle frasi abbiano un impatto sulle prestazioni dei lettori con deficit linguistici o cognitivi; frasi brevi e parole brevi risultano più leggibili non solo ai lettori in generale, ma anche a persone con disabilità cognitive.

I risultati ottenuti da 2.500 lavoratori mostrano che gli utenti del sistema di Crowdsourcing possono effettivamente valutare i livelli di semplicità. Questo metodo di valutazione può quindi essere utile per la valutazione del testo semplificato, rispetto alle

misure standard che classificano i risultati in categorie: “semplificato” o “non semplificato”.

Il loro studio fornisce la prova che il Crowdsourcing è un metodo utile per valutare la semplificazione del testo in inglese e potrebbe anche avere altre applicazioni in futuro, come quella di essere utilizzato per creare il testo semplificato.

La conclusione principale di questo studio è che il pubblico non esperto reclutato da Amazon Mechanical Turk può percepire e misurare i diversi livelli di semplicità nel testo, dimostrando così che il Crowdsourcing può essere uno strumento efficace per valutare l’accessibilità del testo.

2.3 L’approccio ibrido: un esempio per la lingua spagnola

Horacio Saggion, Sanja Stajner, Stefan Bott, Simon Mille, Luz Rello e Biljana Drndarevic hanno svolto una ricerca per riuscire ad ottenere testi più semplici ed accessibili per lettori affetti da bisogni speciali.

Al giorno d’oggi grazie ad internet abbiamo la possibilità di accedere facilmente a molte più informazioni che in qualsiasi altro momento della storia, il problema è che molte di queste informazioni possono essere scritte in uno stile che può rendere il discorso ed i contenuti difficili da capire. Questo può avere un impatto negativo su persone con un livello di alfabetizzazione basso o affette da disturbi cognitivi; purtroppo la consapevolezza di questo è ancora molto bassa e la maggior parte delle informazioni in internet viene pubblicata non tenendo conto delle esigenze di queste particolari fasce di lettori.

Un modo per affrontare questo problema è quello di adattare manualmente il contenuto testuale esistente per renderlo accessibile a queste categorie di utenti, questo tipo di approccio però è molto costoso, sia intermini di tempo che di risorse. Un possibile rimedio è quello di utilizzare tecniche di semplificazione automatica del testo per ottenere testi accessibili a particolari fasce di lettori; il testo semplificato

automaticamente però può subire una importante perdita di informazioni o può risultare troppo semplice o noioso, questa non è necessariamente una cosa negativa se il messaggio del testo originale viene trasmesso al lettore.

L'approccio del gruppo di ricerca spagnolo alla Text Simplification è stato spinto dalla consapevolezza dell'esistenza di questo problema e dalla voglia di trovare una soluzione veloce ed efficace.

2.3.1 Fase preliminare: costruzione e annotazione di un corpus parallelo monolingue

Il gruppo di ricerca spagnolo, all'inizio della loro ricerca, hanno compilato un Corpus composto da 200 articoli di notizie originali semplificati manualmente in lingua spagnola, che gli sono stati forniti dall'agenzia spagnola Servimedia¹. La semplificazione dei testi è stata fatta dai redattori tenendo conto delle esigenze particolari del loro target di utenti di destinazione (persone con disabilità cognitive) seguendo una serie di semplici linee guida.

Il Corpus è stato esaminato al fine di individuare i diversi tipi di operazioni di semplificazione utilizzati e, in seguito, preparare la loro possibile implementazione computazionale. I cambiamenti osservati per quanto riguarda la semplificazione del Corpus possono in gran parte essere raggruppati come segue:

1. **Operazioni sintattiche:** si tratta di modifiche applicate a livello di frase, come lo split della frase o il suo riordinamento;
2. **Operazioni lessicali:** si tratta di sostituire termini poco frequenti, lunghi o tecnici con i loro sinonimi più semplici, di parafrasare alcune espressioni o modificarle;

¹ <http://www.servimedia.es/>

3. **Riduzione del contenuto:** si tratta dell'eliminazione di una porzione significativa del contenuto originale, in conformità con le linee guida che indicano che solo la parte contenente le informazioni essenziali dovrebbe essere conservata;
4. **Precisazione:** si tratta di spiegare dandone una definizione quei termini e concetti complessi per i quali non si riescono a trovare sinonimi.

2.3.2 Fase di valutazione della Text Simplification: misure qualitative ed automatiche

Il modo ideale di valutare i risultati ottenuti dalla semplificazione automatica del testo utilizzata per ottenere testi accessibili ad un certo target della popolazione sarebbe quello di testare la sua efficacia, sui tempi di lettura e sulla comprensione, direttamente rivolgendosi a chi fa parte di quel target; tuttavia interpellare questo target può essere difficile, per questo solitamente per valutare la Text Simplification vengono interrogati utenti (non gli utenti finali) che forniscono punteggi per quanto riguarda la grammaticalità; dato che tale valutazione viene eseguita solo a livello di frase di solito è combinata con la valutazione automatica della semplicità dell'intero testo, misurata in termini di leggibilità. Per quanto riguarda la valutazione degli utenti i possibili parametri di valutazione sono: la grammaticalità, la scorrevolezza del discorso, la conservazione dell'informatività (adeguatezza) e la semplicità.

Nel caso in esame, per la valutazione sono state utilizzate misure automatiche di leggibilità e la valutazione umana. La ricerca puntava ad indagare in che misura l'esito delle semplificazioni si può valutare in termini di metriche automatiche, misurando la complessità del testo in base a diversi livelli linguistici e vedere fino a che punto questi parametri automatici siano correlati ai giudizi dei lettori umani.

Per la valutazione con l'uso di metriche, i ricercatori hanno confrontato prima i testi originali con i testi semplificati automaticamente per vedere se ci sono miglioramenti da

osservare; contemporaneamente hanno confrontato i testi originali con le versioni semplificate da soggetti umani. In questo modo hanno notato che le metriche automatiche solitamente possono essere applicate solo a livello del testo e non a livello della frase, al contrario della valutazione umana che invece è stata progettata per essere applicata a livello della frase.

Per quanto riguarda la valutazione con soggetti umani hanno solo confrontato l'originale con i risultati prodotti dal sistema Simplext che applica solo un gruppo ristretto di trasformazioni basate su regole linguistiche (le semplificazioni fatte dagli umani non sono state valutate).

3. IL CORPUS

La strategia adottata in questo lavoro per testare strumenti per la text simplification è simile a quella spagnola. Questo metodo necessita di un corpus sul quale testare gli strumenti. In questo capitolo introdurremo il corpus costruito a questo scopo, partendo da una definizione preliminare del concetto di corpus.

Un corpus è una collezione di testi utilizzata per ricavare informazioni sul linguaggio. Creare e progettare un corpus significa selezionare ed organizzare i testi secondo precisi criteri, allo scopo di fornire un campione rappresentativo delle varietà e delle tendenze linguistiche di una specifica popolazione linguistica che si intende analizzare. “corpus” infatti può riferirsi a qualsiasi raccolta strutturata di testi, ma oggi questo termine è utilizzato solo in riferimento a raccolte che sono state digitalizzate.

L'avvento dell'era informatica ha rivoluzionato anche l'uso stesso dei corpora, in quanto il computer oggi permette di immagazzinare quantità di dati testuali prima inimmaginabili, creando numerosi vantaggi fra cui un notevole miglioramento dell'accessibilità ai dati. Grazie all'emergere dei linguaggi standard di marcatura del testo, come ad esempio XML², inoltre è stata resa possibile l'annotazione³ linguistica su ampia scala: un testo riccamente annotato costituisce un ottimo filtro di ricerca che ci permette, grazie all'utilizzo di strumenti informatici specializzati, di interrogare in maniera avanzata e rapida il contenuto del corpus.

² Extensible Markup Language (XML) è un linguaggio di markup che definisce un insieme di regole per codificare i testi in un formato leggibile per l'uomo e la macchina.

³ “Annotare” un testo significa rendere esplicite informazioni (meta-)linguistiche mediante l'attribuzione di una etichetta a una porzione specifica del testo.

3.1 Il Corpus del progetto

Quando si cerca di creare un corpus per lo sviluppo di sistemi per Automatic Text Simplification è necessario innanzitutto capire la costituzione del target al quale il sistema dovrà fornire come risultato un testo semplificato. Ovviamente al fine di ottenere una corretta semplificazione automatica non basta saper individuare il target di riferimento e i testi più adatti a quel target, questo è però un primo passo importante.

Nel nostro caso abbiamo selezionato come target studenti italiani di ogni ordine e grado, prendendo come testi di riferimento gli INVALSI ed altri testi di verifica dell'apprendimento; questi testi sono stati scelti perché, oltre a essere classificati per classe, sono tutti seguiti da domande relative alle informazioni contenute nel testo. Queste domande ci serviranno quindi per valutare, se dopo la semplificazione del testo, i testi semplificati hanno mantenuto la loro informatività o l'hanno persa.

3.1.1 INVALSI e altri testi

L'INVALSI è un Ente di ricerca, dotato di personalità giuridica, incentrato sugli aspetti valutativi e qualitativi del sistema scolastico; effettua infatti verifiche periodiche e sistematiche sulle conoscenze e le abilità degli studenti e sulla qualità complessiva dell'offerta formativa delle istituzioni, studia le cause dell'insuccesso e molto altro.

Nel nostro caso le prove INVALSI ci hanno fornito testi idonei alle seguenti classi: seconda, quarta e quinta elementare; prima e terza media; prima, seconda e terza superiore. Queste prove hanno grande importanza, perché al loro interno contengono domande relative alla comprensione del testo, che ci saranno utili per l'analisi dei testi semplificati.

Gli altri testi presi in considerazione per il nostro Corpus sono stati presi da prove scolastiche di diverse classi reperibili online su diversi siti; la particolarità di questi testi, come quella dei testi INVALSI è quella di avere associate delle domande per la comprensione del testo.

I testi del corpus, divisi per classe scolastica, sono stati separati dalle domande in modo da poter essere analizzati.

3.2 Descrizione del Corpus

I testi, separati dalle relative domande, sono stati divisi in cartelle di cui successivamente durante l'analisi quantitativa verrà mostrata la composizione, anche tramite grafici, per dare meglio l'idea della quantità di testi e di frasi analizzate; verrà inoltre descritto in modo approfondito il tipo di domande preso in considerazione per la nostra analisi, anche tramite esempi.

Verrà fatta l'analisi della complessità linguistica dei testi tramite READ-IT. L'analisi di questi testi, non ancora semplificati, ci serve, inizialmente, per avere un'idea della loro difficoltà, ed in seguito, ci servirà per confrontare la complessità di partenza con quella che verrà calcolata dopo la semplificazione dei testi.

L'analisi e la valutazione delle domande è stata fatta manualmente, prendendo un testo alla volta e controllando le domande ad esso associate, alcuni tipi di domande, come vedremo sono state scartate, e le restanti domande sono state divise in tre gruppi. Manualmente è stata anche fatta una prima analisi quantitativa sul numero di testi appartenente ad ogni specifica classe scolastica ed al numero di testi in generale.

Per quanto riguarda invece l'analisi della complessità dei testi, il numero di frasi analizzate ed il loro grado di difficoltà, siamo dovuti ricorrere ad un metodo di analisi automatica, utilizzando uno strumento chiamato READ-IT.

3.2.1 Analisi quantitativa del Corpus

Il corpus è composto da 137 testi, ognuno dei quali ha il proprio file di riferimento contenente le domande necessarie per la comprensione del testo.

I 137 testi sono stati divisi in cartelle a seconda della loro provenienza e del livello di difficoltà a cui appartengono, dove possibile, sono stati inoltre divisi per classe scolastica:

- **INVALSI:**
 - **Elementari:** seconda, quarta, quinta;
 - **Medie:** prima, terza;
 - **Superiori:** prima, seconda, terza.
- **Schola** contente testi di prove delle scuole elementari presi da <http://share.dschola.it>:
 - **Terza elementare;**
 - **Quarta elementare;**
 - **Quinta elementare.**
- **Italiano C13:** contente testi di prove di terza elementare presi da <http://www.maestrasabry.it>;
- **Grammatica** non abbiamo una divisione per classe ma per livello di difficoltà, i testi sono stati presi da <http://grammaticaitaliana.net>:
 - **Elementari:** Livello facile, Livello medio e Livello avanzato;
 - **Medie:** Livello avanzato.
- **Prove:** contente testi di prove di quarta elementare presi da <http://www.scuolevalnervia.org>;
- **Risorse didattiche:** contente testi delle elementari presi da <http://www.risorseditattiche.net>;
- **Vario** contente testi delle elementari presi da <http://maestradituttounpo.blogspot.it> e da <http://www.alphacentauri.it>:

- **Seconda elementare;**
 - **Terza elementare;**
- **Stranieri** contente testi per studenti non madrelingua presi da <http://italianoperstranieri.loescher.it>:
- **Livello C2;**
 - **Livello C3.**

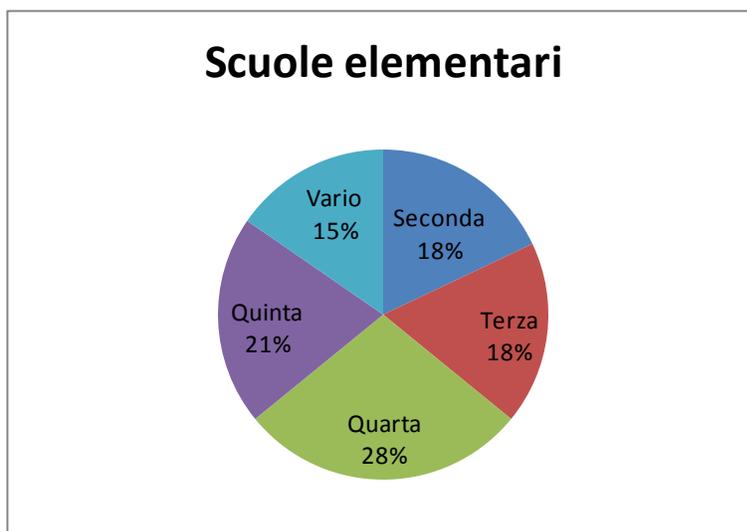


Figura 1: Grafico che mostra in percentuale la quantità di testi analizzati per classe per quanto riguarda i testi delle scuole elementari.

Di questi 137 testi 78 sono testi relativi alle classi delle scuole elementari; in **figura 1** possiamo notare la distribuzione di questi testi per classe.

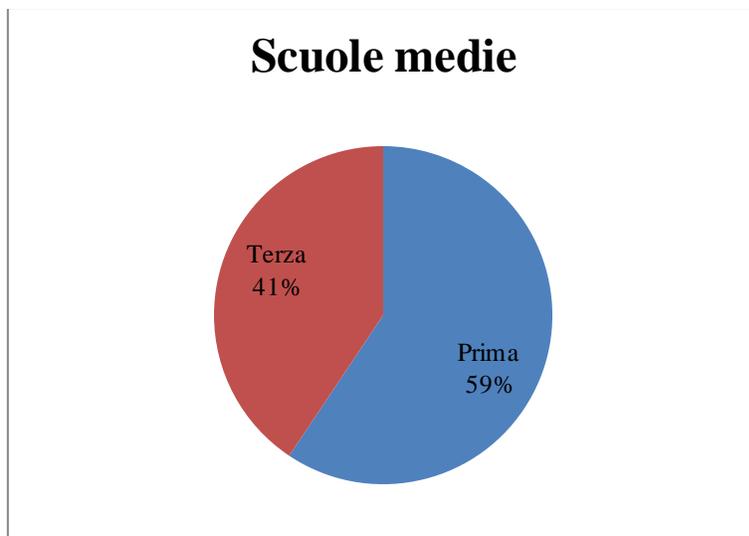


Figura 2: Grafico che mostra in percentuale la quantità di testi analizzati per classe per quanto riguarda i testi delle scuole medie.

Ben 32 testi appartengono alla categoria delle scuole medie, in **figura 2** possiamo vedere la loro distribuzione in base alla classe.

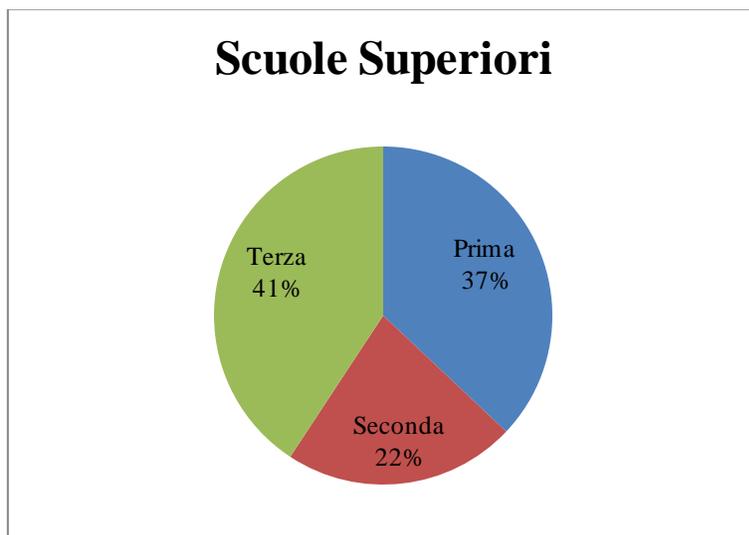


Figura 3: Grafico che mostra in percentuale la quantità di testi analizzati per classe per quanto riguarda i testi delle scuole superiori.

Infine abbiamo 27 testi appartenenti alle scuole superiori, di questi possiamo vedere la loro distribuzione in base alla classe nella **figura 3**.

Questo lavoro di divisione per classe è servito per darci un'idea della complessità dei testi prima di passare alla fase successiva, l'analisi dei testi utilizzando READ-IT; questo perché già in base alla classe scolastica è possibile farsi un'idea della difficoltà di un testo, si parte da un livello di lettura e comprensione semplice per arrivare ad un livello avanzato.

Il corpus contiene ben 3683 frasi, cifra che siamo riusciti a ricavare sommando il numero delle frasi di ogni singolo testo che sono state analizzate da READ-IT. Nella **figura 4** possiamo vedere la distribuzione delle frasi analizzate in base al tipo di scuola (elementari, medie, superiori) e alla classe (prima, seconda, terza, quarta, quinta) non tenendo conto del numero di testi analizzati per categoria.

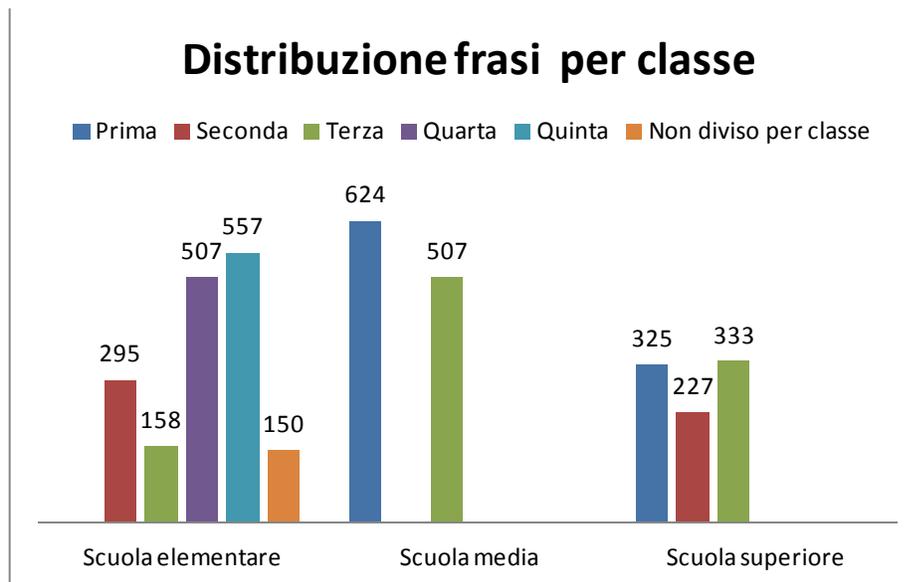


Figura 4: Grafico che mostra il numero delle frasi analizzate da READ-IT divisa per classe scolastica.

Per avere la distribuzione delle frasi tenendo anche conto del numero di testi analizzati per categoria ci basta osservare la **figura 5**, in cui sono riportate le medie delle frasi analizzate per categoria.

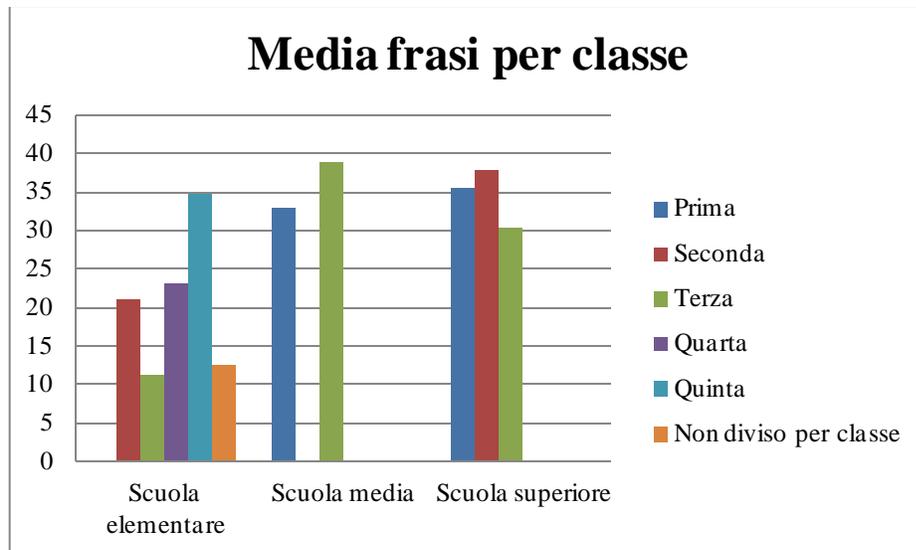


Figura 5: Grafico che mostra la media delle frasi divise per classe e scuola.

Questo grafico tiene conto del numero di testi analizzati per categoria e ci da modo, per esempio, di osservare che nonostante il numero di frasi della seconda superiore sia relativamente basso (**figura 4**), tenendo conto del numero di testi analizzati, possiamo notare che nonostante i testi appartenenti a questa categoria siano pochi (sono 6), la media delle frasi analizzate è alta; questo ci fa capire che i testi appartenenti a questa categoria sono lunghi. Se invece prendiamo ad esempio i testi di seconda elementare, di cui risultano essere state individuate più frasi di quelle analizzate nei testi di seconda superiore (**figura 4**), nella **figura 5** possiamo notare che la distribuzione della media delle frasi è più bassa, questo si spiega perché il numero di testi di seconda elementare è maggiore (sono infatti 14) mentre la lunghezza dei testi è minore.

Durante l'analisi dei testi appartenenti al corpus abbiamo osservato che la lunghezza dei testi aumenta gradualmente con l'avanzare delle classi scolastiche ed allo stesso modo aumenta anche la difficoltà dei testi analizzati.

3.2.2 Le domande e l'analisi del testo

Contemporaneamente all'analisi dei testi tramite READ-IT è stato fatto un altro tipo di analisi del testo per valutare se le domande di comprensione associate ai testi fossero tutte utili ai fini della valutazione dei testi che avverrà successivamente dopo la semplificazione. Alcuni tipi di domande, basate sulla grammatica, sono state scartate come anche le domande in cui veniva richiesto di continuare la storia; questo perché a noi interessava avere, per ogni testo, una o più domande che ci aiutassero a valutare l'informatività del testo.

Le domande relative all'informatività del testo che sono state individuate risultano di tre tipi:

- 1) Domande aperte la cui risposta è contenuta nel testo in modo esplicito;
- 2) Domande con un elenco di possibili risposte, la cui risposta è contenuta nel testo;
- 3) Domande aperte la cui risposta è deducibile dallo studente, ma non presente in modo esplicito nel testo ed ha bisogno di una maggiore elaborazione.

Per capire meglio i diversi tipi di domande possiamo fornirne alcuni esempi:

- 1) *Molti clienti degli alberghi sono superstiziosi. Quali sono i due comportamenti con cui lo manifestano?*

La risposta la troviamo direttamente nel testo, espressa in modo molto chiaro:

“ *All'albergo, se gli propongono la camera numero 13, queste persone la rifiutano e chiedono di essere messe in un'altra camera. Rifiutano addirittura di dormire in qualsiasi camera che si trovi al tredicesimo piano - tant'è vero che, se ci badate, vi accorgete che in tanti alberghi il tredicesimo piano non esiste nemmeno.* “

Questa domanda è tratta da una prova INVALSI di prima media, la prossima è ancora più esaustiva nel mostrare come la risposta sia già scritta esplicitamente nel testo:

1.1) *Di che cosa ha bisogno la mamma?*

La risposta la troviamo direttamente nel testo espressa in modo chiaro ed esplicito:

“ *La mamma ha bisogno subito di un limone e Riccardo...* ”

2) *Secondo l'anatra-capo, che cosa serve per resistere al vento del mare?*

- *Sapere starnazzare*
- *Avere penne e ossa*
- *Sbattere velocemente le ali*
- *Ubbidire al capo*

La risposta la troviamo direttamente nel testo:

“ *- Non ce la farai mai, - gli disse l'anatra-capo guardandolo con disprezzo.*
- Non hai penne, non hai ossa, non ce la farai mai contro il vento del mare, ti farà a brandelli. ”

Questa domanda è relativa ad una prova INVALSI della seconda elementare.

3) *Come si comporta il bambino per strada?*

La risposta non è espressa in modo chiaro ed esplicito nel testo, ma la si può ricavare dopo un'attenta lettura di questa parte del testo:

“ *Riccardo ha otto anni e non è mai andato da solo a fare una commissione. Eccolo sul portone, guarda di qua e di là un momento, poi...via, ben vicino al muro come ha raccomandato la mamma. Passa un autobus, poi un altro. Riccardo può finalmente attraversare.* “

Questa domanda è relativa ad una prova contenuta nella cartella ItalianoC13, per studenti di terza elementare.

3.1) *Quale rapporto si stabilisce tra loro? (bambino e delfino)*

Anche in questo caso la risposta non è espressa in modo chiaro ed esplicito nel testo, ma la si può ricavare dopo la lettura di questa parte del testo:

“ *Il bambino era pigro e il delfino si divertiva a svegliarlo ogni volta che si addormentava sulla sdraio: emergeva dall'acqua e gridava. E quando il bambino era sveglio si tuffava e si rituffava, con aria di dire “non mi acchiappi, non mi acchiappi”. Ma il bambino era sempre imbronciato. Non parlava, non rideva. Per attirare l'attenzione del bambino, il delfino cominciò a fare tuffi, salti e capriole come un delfino da circo. Il primo giorno il bambino scoppiò a piangere, il secondo giorno scoppiò a ridere, il terzo si mise a parlare; il delfino gli rispose con un piccolo grido. Si capirono perfettamente. Il bambino diceva parole e il delfino emetteva piccoli gridi. Si raccontarono molte storie: il bambino raccontò di un gatto, di un orsacchiotto, di una palla e di una trottola; Il delfino raccontò del pescecane, dei coralli e dei pesci-luna.* “

Questa domanda è relativa ad una prova contenuta nella cartella Prove.

3.2) *Come trasformava le bistecche?*

Questa domanda è addirittura a trabocchetto, in quanto nel testo non si parla di bistecche; la possiamo trovare nella cartella Vario. Anche in questo caso possiamo osservare come questo tipo di domande sia diverso dai tipi precedenti.

Analizzando le domande per le scuole elementari delle prove INVALSI possiamo osservare che tutte sono domande con opzioni di risposta. Appartengono quindi al secondo tipo di domande appena analizzato. Per quanto riguarda le prove INVALSI delle medie, invece abbiamo sempre una predominanza di domande di tipo due, ma sono presenti, anche se in minima parte domande del tipo uno. Gli stessi tipi di domande delle scuole medie sono presenti anche nelle prove INVALSI appartenenti alle scuole superiori.

Continuando l'analisi delle domande abbiamo visto che anche quelle appartenenti alla cartella Grammatica e alla cartella Stranieri sono tutte domande del secondo tipo; per quanto riguarda le domande dei testi della cartella ItalianoC13, di Prove, di Schola e di Vario, troviamo domande di tipo uno, di tipo due e di tipo tre; le domande della cartella RisorseDidattiche sono di tipo uno e tre.

3.2.3 Analisi della complessità linguistica

Grazie all'analisi compiuta da READ-IT siamo stati in grado di ottenere diversi livelli di complessità per ogni testo analizzato e per ogni frase di quel testo. In seguito verrà descritto lo strumento utilizzato per ottenere questi risultati e verranno analizzati i suoi diversi livelli di difficoltà, che sono: Base, lessicale, sintattico e globale.

3.2.3.1 READ-IT

READ-IT rappresenta il primo strumento di valutazione della leggibilità per quanto riguarda la lingua italiana. Proprio per questo noi lo utilizzeremo per valutare la complessità linguistica dei testi del nostro corpus prima e dopo la semplificazione.

Il passo preliminare per l'attribuzione di un punteggio di leggibilità è l'annotazione linguistica automatica del testo.

In questo processo il testo in input viene suddiviso in frasi ed ogni frase viene divisa in token attraverso il processo di tokenizzazione. Viene in seguito effettuato il PoS-Tagging, l'annotazione dei tratti morfo-sintattici e insieme l'annotazione sintattica a dipendenze. Viene poi proposta l'analisi della leggibilità, che è quella che a noi interessa.

Testo da analizzare	Suddivisione in frasi	Suddivisione in token	Parti del discorso	Annotazione	Analisi globale della leggibilità	Proiezione della leggibilità sul testo	
indice di leggibilità		livello di difficoltà					
READ-IT Base						15,5%	
READ-IT Lessicale						93,7%	
READ-IT Sintattico						4,7%	
READ-IT Globale						36,7%	

Figura 6: Sezione dell'analisi della leggibilità, divisa nei 4 approcci di valutazione

I diversi indici di leggibilità di READ-IT, come mostra la **Figura 6**, sono:

- **Base:** il testo viene valutato sulla base della lunghezza delle frasi, calcolata come numero medio di parole per frase, e della lunghezza delle parole, calcolata come numero medio di caratteri per parola;
- **Lessicale:** il testo viene valutato focalizzandosi sulle caratteristiche lessicali del testo, come ad esempio la composizione del vocabolario e dalla sua ricchezza lessicale;

- **Sintattico**: il testo viene valutato basandosi su informazioni di tipo grammaticale, ovvero sulla combinazione di tratti morfo-sintattici e sintattici desunti dai corrispondenti livelli di analisi linguistica;
- **Globale**: il testo viene valutato basandosi sulla combinazione di tratti di varia natura, che spaziano dalle caratteristiche generali del testo a quelle lessicali e sintattiche utilizzate, servendosi quindi di tutti e tre gli indici elencati sopra.

Nel caso della **figura 6**, i dati relativi ai livelli di difficoltà del testo appartengono al testo nel suo complesso, sono quindi punteggi globali di complessità. READ-IT ci permette anche di analizzare il testo frase per frase ottenendo così, per ogni frase, il rispettivo livello di complessità.

Testo da analizzare	Suddivisione in frasi	Suddivisione in token	Parti del discorso	Annotazione	Analisi globale della leggibilità	Proiezione della leggibilità sul testo			
						SID	frase	base	less.
1.	Di solito i leoni sono animali feroci, Leo invece ama tutti quanti e gioca dal mattino alla sera con i suoi amici.								
2.	Mamma leonessa è preoccupata e un giorno gli dice: - Se non te la senti di fare il cacciatore devi andartene di qui, non c'è posto per te nel branco.								

Figura 7: Sezione di READ-IT sulla proiezione della leggibilità sul testo.

READ-IT mostra anche la proiezione della leggibilità sul testo, ovvero i valori degli indici di leggibilità calcolati sulle singole frasi. Il colore delle celle delle ultime quattro colonne della tabella nella **figura 7** indica la leggibilità di ogni frase. Il livello di difficoltà è rappresentato cromaticamente mediante colori che vanno dal verde (testo leggibile) al rosso (testo particolarmente difficile).

Per ottenere dei valori numerici riguardo al grado di leggibilità basta passare col mouse su ogni singola cella colorata per visualizzarne il valore sottoforma di percentuale. Le frasi prese in analisi in **figura 7** appartengono ad uno dei testi del nostro corpus che sono stati analizzati, di seguito riporteremo i valori numerici relativi ai punteggi dei diversi indici:

Frase	Base	Lessicale	Sintattico	Globale
1	41,4%	39,3%	36,0%	7,3%
2	63,0%	39,9%	14,8%	95,8%

Un eventuale utente che utilizza questo strumento per la valutazione di un testo può affidarsi alle diverse colorazioni per valutare se una frase va semplificata o meno e da che punto di vista (lessico o sintattico) va effettuata questa semplificazione.

3.2.3.2 Risultati

Ora per ogni classe analizzeremo i risultati ottenuti dall'analisi con READ-IT:

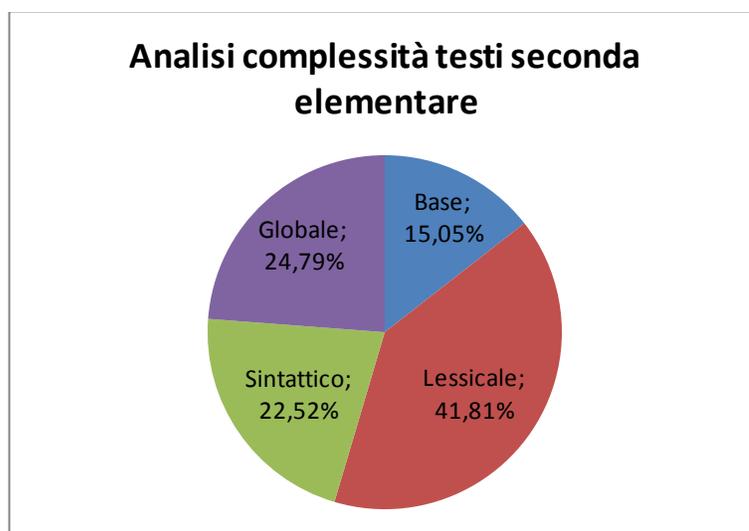


Figura 8: Grafico che riporta la percentuale della complessità dei testi appartenenti alla seconda elementare.

Per quanto riguarda la seconda elementare, come possiamo osservare in **figura 8**, la complessità globale del testo non è particolarmente elevata, come non lo sono quella sintattica e quella di base, mentre è abbastanza elevata quella lessicale.

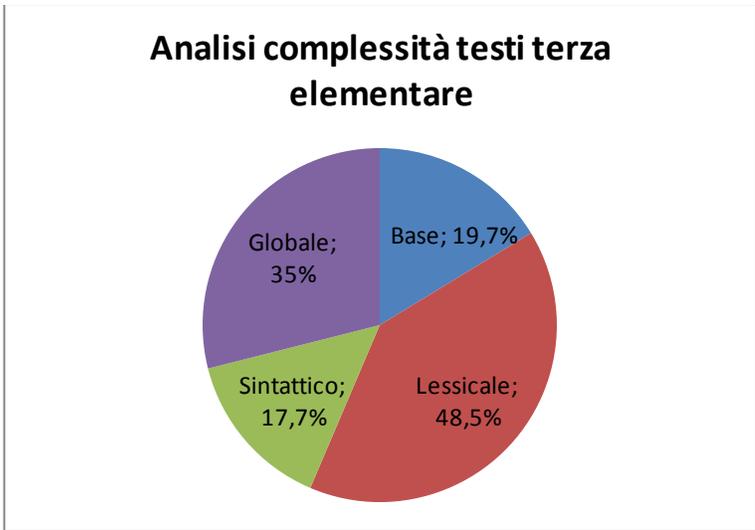


Figura 9: Grafico che riporta la percentuale della complessità dei testi appartenenti alla terza elementare.

Per quanto riguarda la terza elementare, come possiamo osservare in **figura 9**, la complessità globale non è molto elevata, come non lo sono quella sintattica e quella di base, continua invece ad essere abbastanza elevata la complessità lessicale.

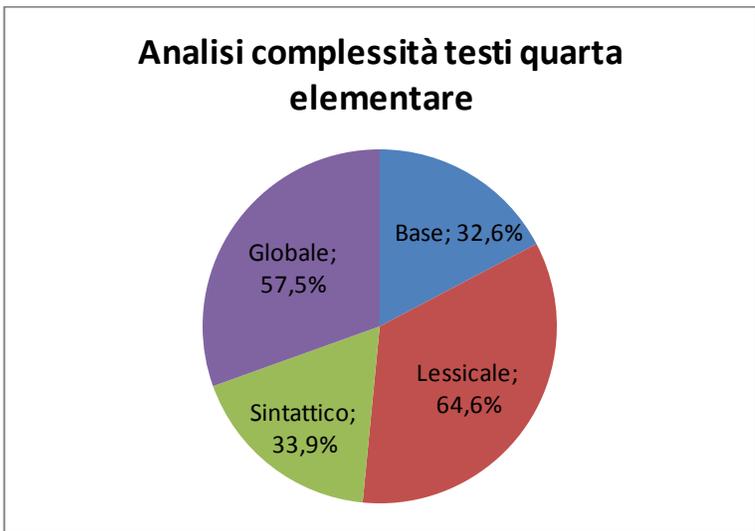


Figura 10: Grafico che riporta la percentuale della complessità dei testi appartenenti alla quarta elementare.

Per quanto riguarda la quarta elementare, come possiamo osservare in **figura 10**, la complessità globale inizia ad essere più alta che nelle precedenti, sono aumentate anche la sintattica e quella di base, ma la più alta rimane sempre quella lessicale.

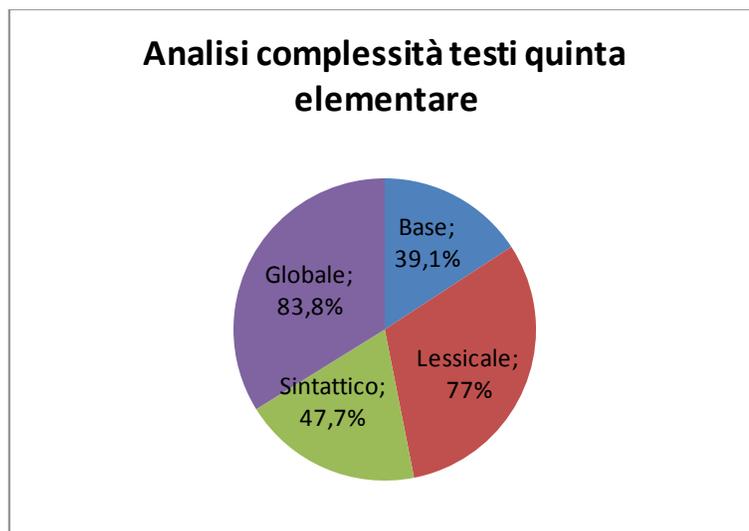


Figura 11: Grafico che riporta la percentuale della complessità dei testi appartenenti alla quinta elementare.

Per quanto riguarda la quinta elementare, come possiamo osservare in **figura 11**, la complessità globale è elevata, supera persino quella lessicale, che nelle analisi della classi precedenti non ha mai superato. Sono aumentate anche la sintattica e quella di base.

Analisi complessità testi elementari non classificati per classe

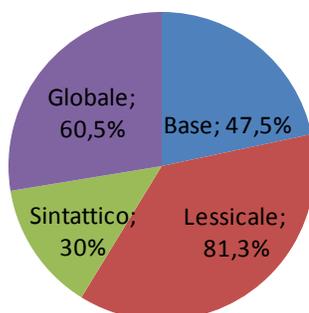


Figura 12: Grafico che riporta la percentuale della complessità dei testi appartenenti alla scuola elementare che non sono stati classificati per classe.

In **figura 12**, possiamo osservare la complessità di testi di cui non abbiamo una divisione predefinita per classe scolastica, ma che sappiamo appartengono alla scuola elementare.

Analisi complessità testi prima media

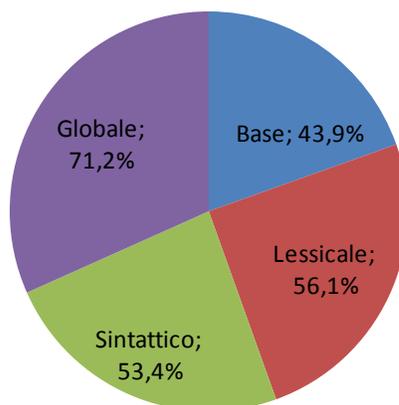


Figura 13: Grafico che riporta la percentuale della complessità dei testi appartenenti alla prima media.

Per quanto riguarda la prima media, come possiamo osservare in **figura 13**, la complessità globale è elevata più degli altri livelli di complessità analizzati.

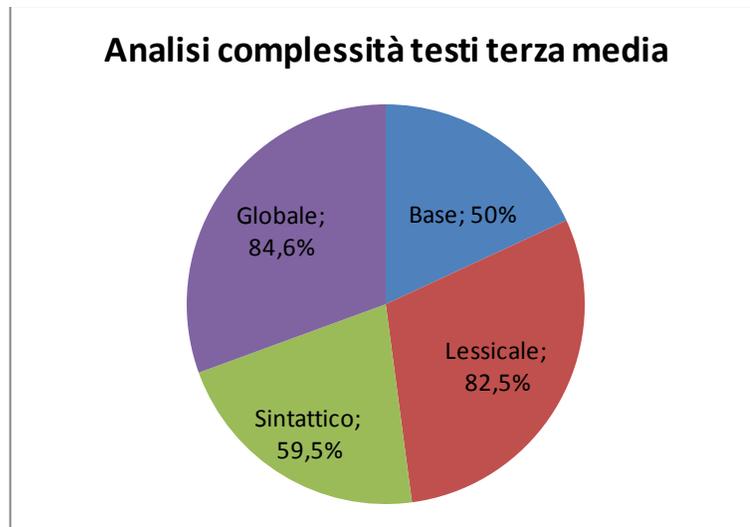


Figura 14: Grafico che riporta la percentuale della complessità dei testi appartenenti alla terza media.

Per quanto riguarda la terza media, come possiamo osservare in **figura 14**, la complessità globale è elevata più degli altri livelli di complessità analizzata e in confronto a quanto mostra la **figura 13** la complessità lessicale, la sintattica e quella di base sono più elevate.

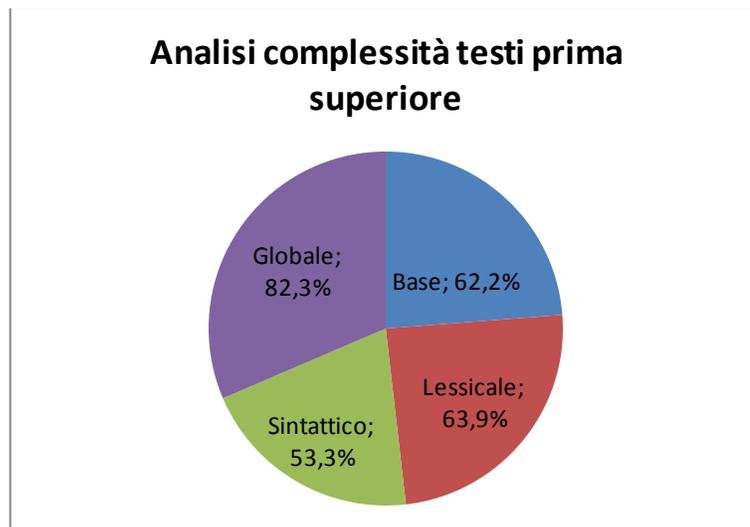


Figura 15: Grafico che riporta la percentuale della complessità dei testi appartenenti alla prima superiore.

Per quanto riguarda la prima superiore, come possiamo osservare in **figura 15**, la complessità globale è elevata più degli altri livelli di complessità analizzati.

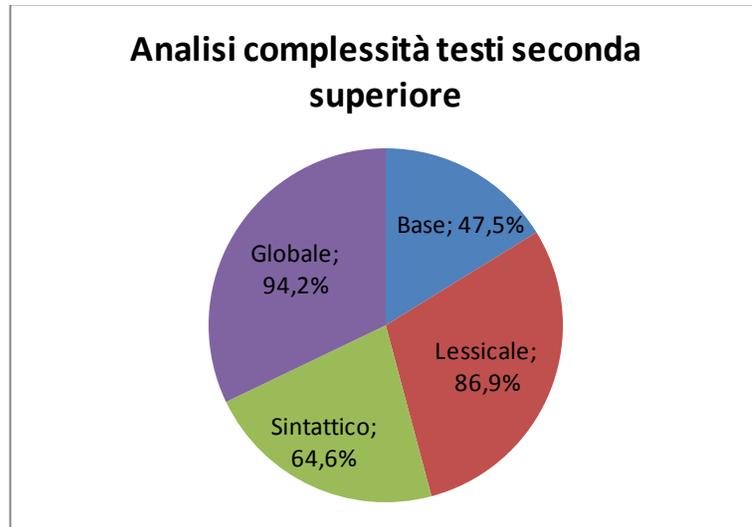


Figura 16: Grafico che riporta la percentuale della complessità dei testi appartenenti alla seconda superiore.

Per quanto riguarda la seconda superiore, come possiamo osservare in **figura 16**, la complessità globale, quella sintattica e quella lessicale sono più elevate che in **figura 15**, tuttavia il livello di complessità di base è inferiore.

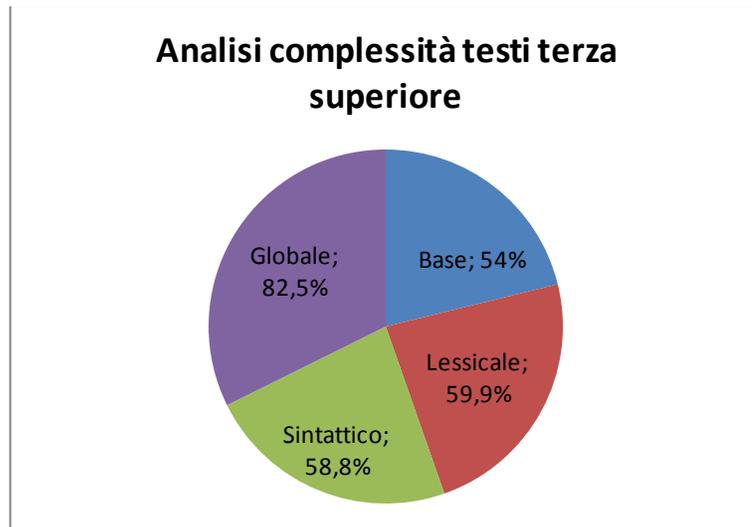


Figura 17: Grafico che riporta la percentuale della complessità dei testi appartenenti alla terza superiore.

Per quanto riguarda la terza superiore, come possiamo osservare in **figura 17**, la complessità globale, la lessicale e la sintattica sono inferiori a quelli riportati in **figura 16**, tuttavia la complessità di base è superiore.

Grazie a questi grafici possiamo osservare come i livelli di complessità, nella maggior parte dei casi aumentino con l'aumentare della classe scolastica. Grazie alla **figura 18** possiamo osservare meglio i livelli di complessità globale di tutte le classi scolastiche prese in analisi.

Il cambiamento di livello di complessità globale è rappresentato in modo evidente per quanto riguarda i testi delle scuole elementari e delle scuole medie. Questo cambiamento non è ben rappresentato per quanto riguarda i testi delle scuole superiori perché i testi delle superiori hanno un livello di difficoltà molto più uniforme in confronto ai testi delle scuole precedenti.

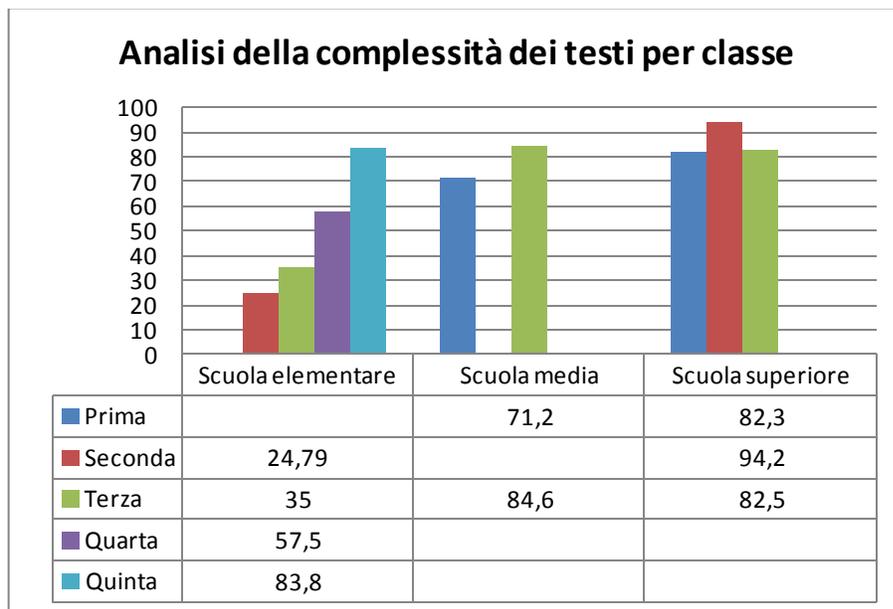


Figura 18: Grafico che rappresenta la media dei risultati dell'analisi della complessità globale per classe.

4. ANALISI DEL TESTO SEMPLIFICATO

In questo capitolo parleremo dei risultati ottenuti dopo la semplificazione del testo avvenuta con il DELETE. La semplificazione è il risultato di un sistema automatico che ha eliminato una o più parole dalle frasi dei testi originali. Saranno inoltre riportati grafici che rappresentano la complessità del testo dopo la semplificazione, utilizzando come in precedenza lo strumento READ-IT, i dati relativi alla complessità del testo verranno inoltre confrontati con i dati precedenti alla semplificazione.

Verranno descritti i criteri di valutazione delle frasi semplificate applicati manualmente: per ogni testo semplificato sono state analizzate le frasi per verificare la loro grammaticalità e la loro informatività. Quindi saranno riportati esempi di frasi semplificate che dopo l'analisi avvenuta manualmente sono state ritenute corrette, che hanno cioè mantenuto la grammaticalità e l'informatività, ed esempi di frasi semplificate che invece hanno perso grammaticalità o informatività.

4.1 Analisi dei risultati di READ-IT

I testi una volta sottoposti alla semplificazione con l'applicazione del DELETE sono stati sottoposti nuovamente all'analisi di complessità, tramite READ-IT, per verificare i cambiamenti di complessità ottenuti dopo la semplificazione.

In alcuni casi ci sono stati testi che non hanno subito alcun cambiamento dopo la semplificazione, per questo non sono stati trattati in questa fase, in quanto non sono utili per ottenere risultati relativi alla semplificazione dei testi.

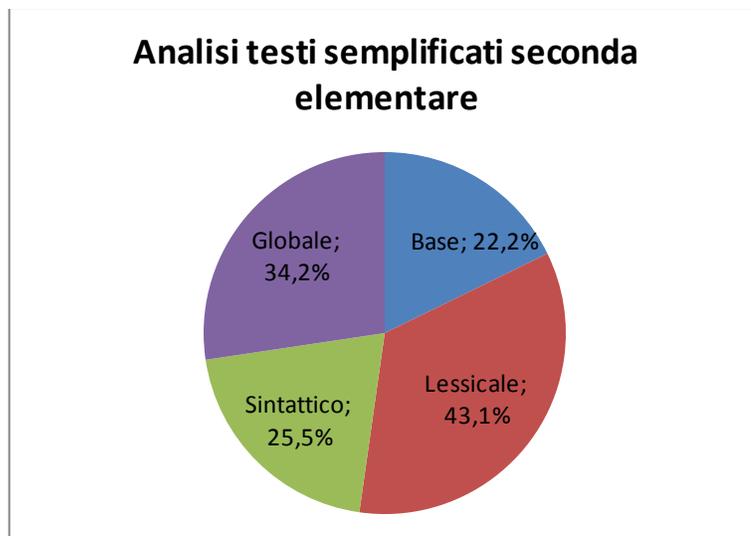


Figura 21: Grafico che riporta la percentuale di complessità dei testi semplificati di seconda elementare.

Per quanto riguarda la complessità dei testi di seconda elementare, la possiamo osservare in **figura 21**. Nel caso dei testi di seconda elementare, un testo non è stato preso in considerazione, in quanto non ha subito cambiamenti rispetto al testo originale.

Anche se, in questo caso un testo non è stato preso in analisi, confrontando questi risultati con i precedenti, riportati in **figura 8**, la complessità dei testi non è diminuita dopo la semplificazione, anzi è aumentata.

Naturalmente abbiamo verificato i risultati escludendo dai dati analizzati in **figura 8** il testo che non ha subito modifiche:

Base	Lessicale	Sintattico	Globale
15,6%	38,0%	23,2%	24,9%

Il risultato finale non cambia, i testi semplificati risultano, più complessi di quelli non semplificati.

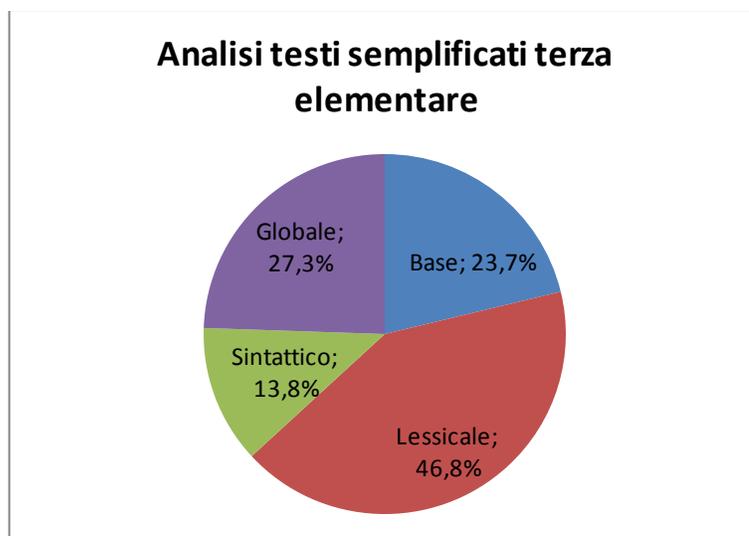


Figura 22: Grafico che riporta la percentuale di complessità dei testi semplificati di terza elementare.

Per quanto riguarda la complessità dei testi semplificati di terza elementare, la possiamo osservare in **figura 22**. Nei testi di terza elementare, ben cinque testi non sono stati presi in considerazione, in quanto non hanno subito cambiamenti rispetto ai testi originali.

Confrontando i risultati riportati in **figura 22** con quelli riportati in **figura 9**, la complessità dei testi ha subito dei cambiamenti rilevanti dopo la semplificazione, possiamo osservare una diminuzione della complessità di tutti i livelli analizzati.

Per avere dei risultati corretti abbiamo escluso dai dati analizzati in **figura 9** i testi che non hanno subito modifiche:

Base	Lessicale	Sintattico	Globale
23,2%	39,9%	15,7%	26,9%

Grazie a questi risultati possiamo osservare una lieve diminuzione del livello di complessità di base, una diminuzione un po' più rilevante della complessità lessicale, e un aumento sia del livello di complessità sintattico sia di quello globale.

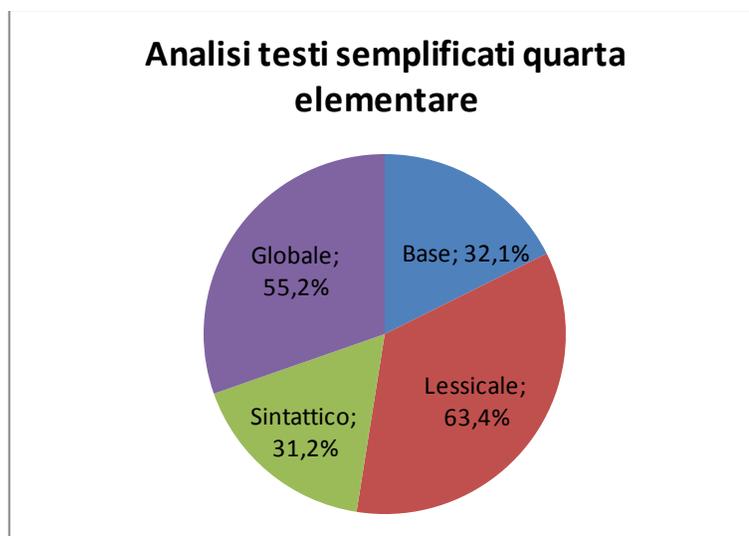


Figura 23: Grafico che riporta la percentuale di complessità dei testi semplificati di quarta elementare.

Per quanto riguarda la complessità dei testi semplificati di quarta elementare, la possiamo osservare in **figura 23**. In questo caso solo un testo non è stato preso in considerazione, in quanto non ha subito cambiamenti rispetto ai testi originali.

Confrontando i risultati riportati in **figura 23** con quelli riportati in **figura 10**, la complessità dei testi ha subito dei cambiamenti rilevanti dopo la semplificazione, possiamo osservare una diminuzione della complessità di tutti i livelli analizzati.

Per avere dei risultati corretti abbiamo escluso dai dati analizzati in **figura 9** i testi che non hanno subito modifiche:

Base	Lessicale	Sintattico	Globale
31,7%	63,1%	32,3%	57,1%

Grazie a questi risultati possiamo osservare una diminuzione del livello di complessità di base e della complessità lessicale, ed un aumento sia del livello di complessità sintattico sia di quello globale.

Analisi testi semplificati quinta elementare

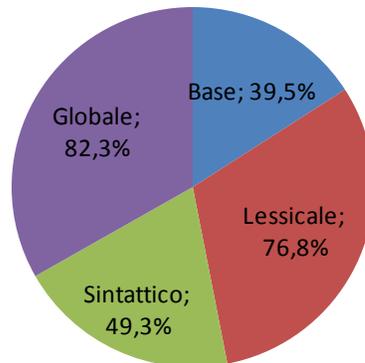


Figura 24: Grafico che riporta la percentuale di complessità dei testi semplificati di quinta elementare.

Per quanto riguarda la complessità dei testi semplificati di quinta elementare, la possiamo osservare in **figura 24**. In questo caso nessun testo è stato escluso dall'analisi, in quanto tutti i testi hanno subito cambiamenti rispetto ai testi originali.

Possiamo quindi osservare confrontando la **figura 24** con la **figura 11** i cambiamenti relativi alla complessità dei testi, possiamo notare che rispetto ai testi originali quelli semplificati hanno: un livello di complessità di base e di complessità sintattica più alto ed un livello di complessità lessicale e globale più basso.

Analisi testi semplificati elementari non classificati per classe

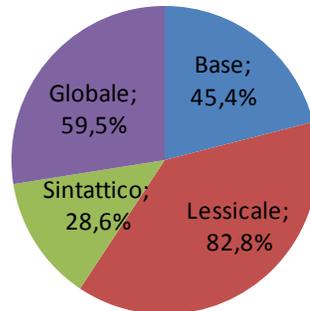


Figura 25: Grafico che riporta la percentuale di complessità dei testi semplificati delle elementari non classificati per classe scolastica.

Per quanto riguarda la complessità dei testi semplificati delle elementari non classificati per classe scolastica, la possiamo osservare in **figura 25**. Anche in questo caso nessun testo è stato escluso dall'analisi, in quanto tutti i testi hanno subito cambiamenti rispetto ai testi originali.

Possiamo quindi osservare confrontando la **figura 25** con la **figura 12** i cambiamenti relativi alla complessità dei testi, possiamo notare che rispetto ai testi originali quelli semplificati hanno un livello di complessità di base, di complessità sintattica e globale più bassi, mentre solo il livello di complessità lessicale è più alto.

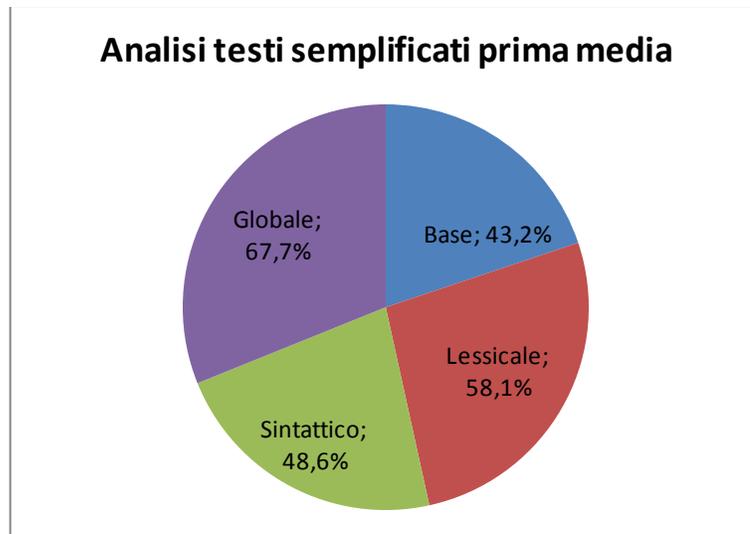


Figura 26: Grafico che riporta la percentuale di complessità dei testi semplificati della prima media.

Per quanto riguarda la complessità dei testi semplificati di prima media, la possiamo osservare in **figura 26**. Anche in questo caso nessun testo è stato escluso dall'analisi, in quanto tutti i testi hanno subito cambiamenti rispetto ai testi originali.

Anche in questo caso, possiamo quindi osservare confrontando la **figura 26** con la **figura 13** i cambiamenti relativi alla complessità dei testi, possiamo notare che rispetto ai testi originali quelli semplificati hanno: un livello di complessità di base e di complessità sintattica e globale più bassi, mentre solo il livello di complessità lessicale è più alto.

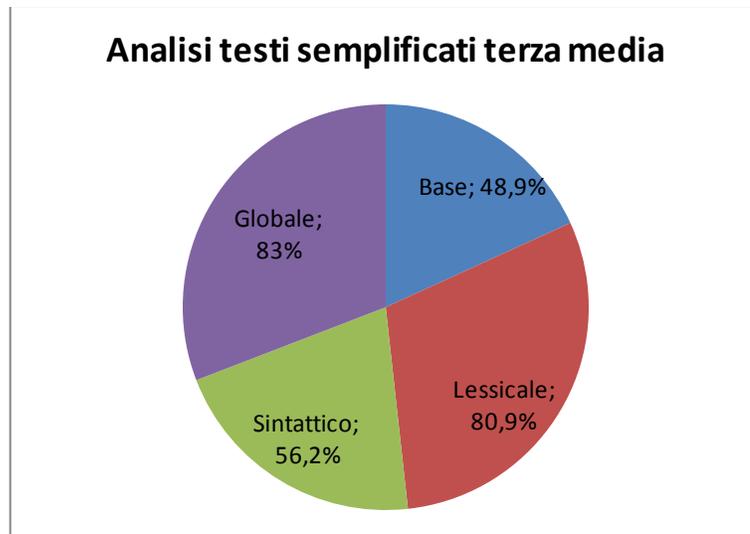


Figura 27: Grafico che riporta la percentuale di complessità dei testi semplificati della terza media.

Per quanto riguarda la complessità dei testi semplificati di terza media, la possiamo osservare in **figura 27**. Anche in questo caso nessun testo è stato escluso dall'analisi, in quanto tutti i testi hanno subito cambiamenti rispetto ai testi originali.

Possiamo quindi osservare confrontando la **figura 27** con la **figura 14** i cambiamenti relativi alla complessità dei testi: tutti i livelli di complessità risultano più bassi.

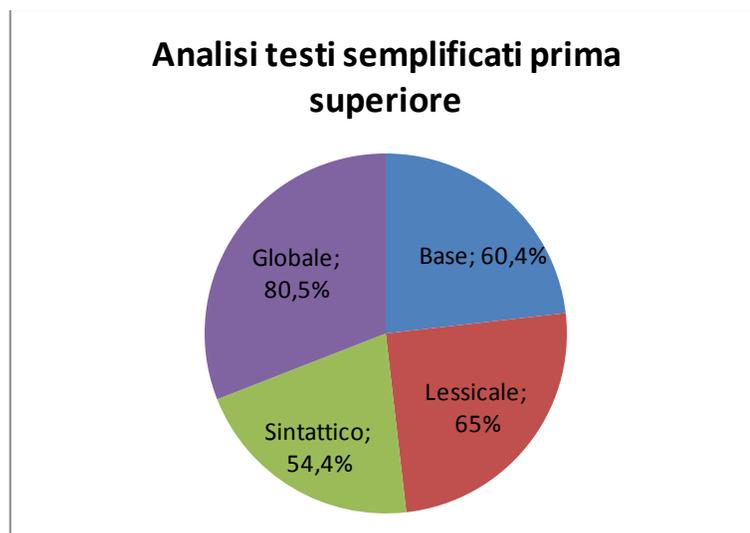


Figura 28: Grafico che riporta la percentuale di complessità dei testi semplificati della prima superiore.

Per quanto riguarda la complessità dei testi semplificati di prima superiore, la possiamo osservare in **figura 28**. Anche in questo caso nessun testo è stato escluso dall'analisi, in quanto tutti i testi hanno subito cambiamenti rispetto ai testi originali.

Possiamo quindi osservare confrontando la **figura 28** con la **figura 15** i cambiamenti relativi alla complessità dei testi, possiamo notare che rispetto ai testi originali quelli semplificati hanno: un livello di complessità di base e di complessità globale più bassi, mentre il livello di complessità lessicale e quello sintattico sono più alti.

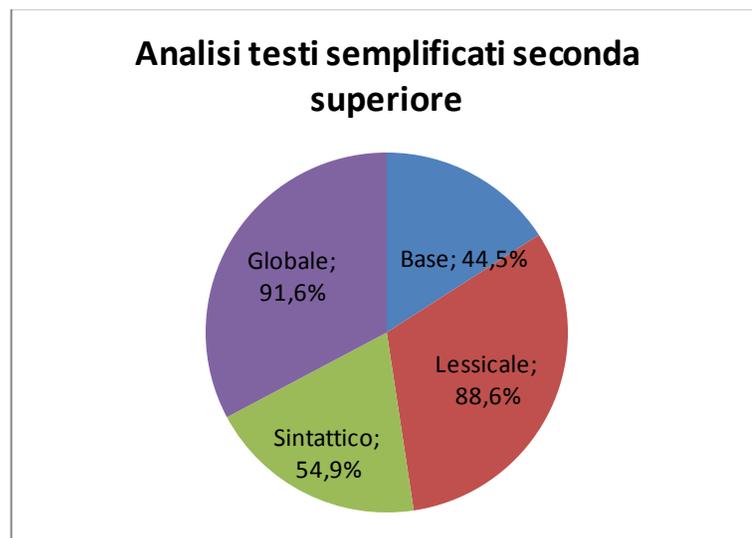


Figura 29: Grafico che riporta la percentuale di complessità dei testi semplificati della seconda superiore.

Per quanto riguarda la complessità dei testi semplificati di seconda superiore, la possiamo osservare in **figura 29**. Anche in questo caso nessun testo è stato escluso dall'analisi, in quanto tutti i testi hanno subito cambiamenti rispetto ai testi originali.

Possiamo quindi osservare confrontando la **figura 29** con la **figura 16** i cambiamenti relativi alla complessità dei testi, possiamo notare che rispetto ai testi originali quelli semplificati hanno: un livello di complessità di base e di complessità sintattica e globale più bassi, mentre solo il livello di complessità lessicale è più alto.

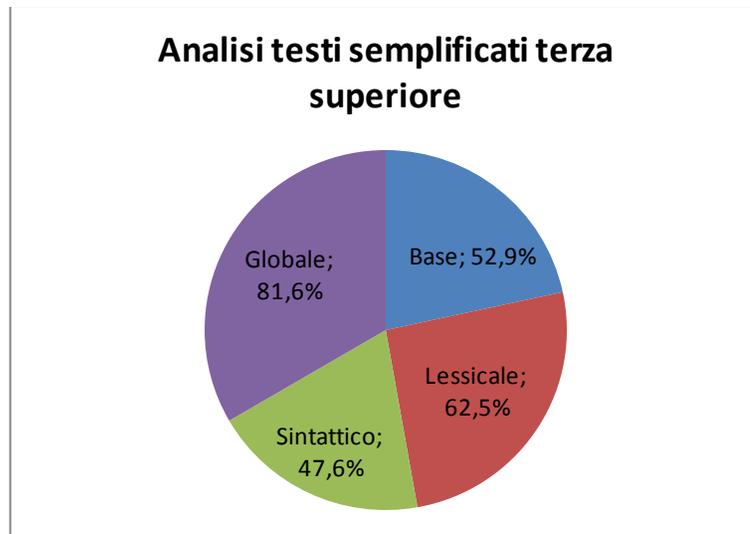


Figura 30: Grafico che riporta la percentuale di complessità dei testi semplificati della terza superiore.

Per quanto riguarda la complessità dei testi semplificati di seconda superiore, la possiamo osservare in **figura 30**. Anche in questo caso nessun testo è stato escluso dall'analisi, in quanto tutti i testi hanno subito cambiamenti rispetto ai testi originali.

Possiamo quindi osservare confrontando la **figura 30** con la **figura 17** i cambiamenti relativi alla complessità dei testi, possiamo notare che rispetto ai testi originali quelli semplificati hanno: un livello di complessità di base e di complessità sintattica e globale più bassi, mentre solo il livello di complessità lessicale è più alto.

Infine possiamo analizzare i risultati nel loro complesso, confrontando la media dei livelli di difficoltà rilevati per tutti i testi prima della semplificazione con quelli calcolati dopo la semplificazione dei testi.

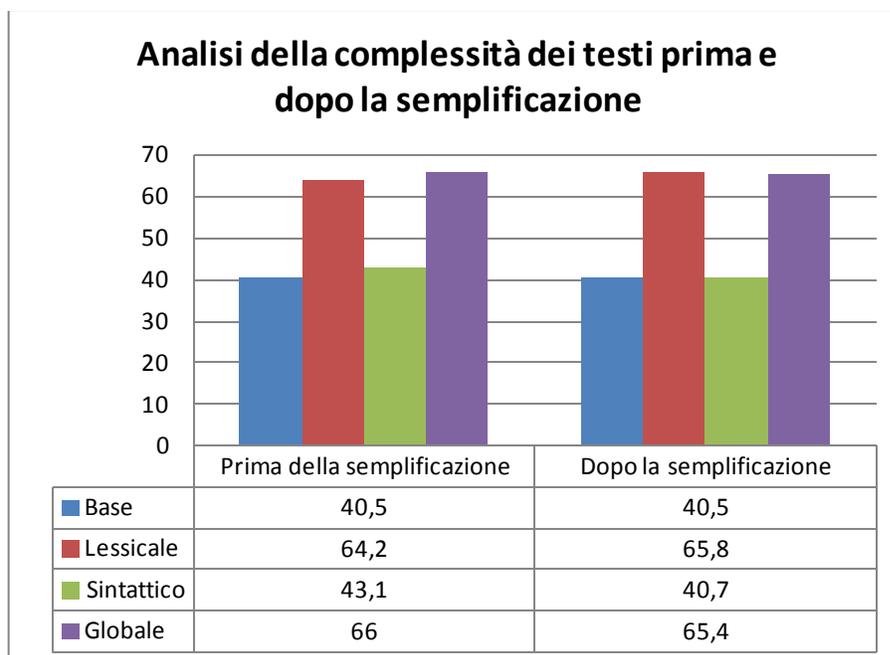


Figura 31: Grafico che riporta la complessità di tutti i testi prima della semplificazione e dopo la semplificazione.

Grazie al grafico di **figura 31** possiamo notare come la semplificazione del testo abbia influito sui diversi livelli di complessità analizzati grazie a READ-IT. Le medie dei valori di complessità dei testi semplificati differiscono dalle medie del testo di partenza di molto poco.

Tra i testi di partenza ed i testi semplificati vi è una differenza quasi impercettibile guardando solo le colonne colorate, è per questo che al grafico a colonne è stata aggiunta una tabella con valori numerici. Grazie a questa tabella possiamo osservare che il livello di complessità di base è rimasto invariato, mentre è aumentato il livello lessicale, il livello sintattico è diminuito ed il livello globale è diminuito in modo quasi impercettibile.

4.2 Analisi dei risultati svolte manualmente

Dopo l'applicazione del DELETE i testi sono stati analizzati in modo automatico tramite READ-IT, come visto sopra, successivamente sono stati analizzati manualmente per verificarne la grammaticalità e l'informatività.

Sono state prese in esame ben 736 frasi, tutte quelle che hanno subito una modifica dopo la semplificazione.

Ora vediamo nel dettaglio com'è avvenuta l'analisi della grammaticalità e quella dell'informatività ed i rispettivi risultati riscontrati.

4.2.1 Analisi della grammaticalità

La verifica della grammaticalità è avvenuta semplicemente leggendo le frasi e verificandone la correttezza grammaticale; in molte frasi che risultano non grammaticali è stata riscontrata l'assenza di articoli dopo la semplificazione, o l'assenza di parole che rendono la frase priva di senso.

Ecco alcuni esempi di frasi che dopo la semplificazioni hanno mantenuto o meno la grammaticalità:

- **Prima:** “*La mamma frugò nella libreria, finchè non trovò un grosso librone rosso.*”

Dopo: “*La mamma frugò nella libreria.*”

In questo caso la frase ha mantenuto la sua grammaticalità anche se dopo la semplificazione, sia stata cancellata l'intera subordinata dopo la virgola.

- **Prima:** “*Questo è ciò che propone un B&B di Quartu Sant'Elena in provincia di Cagliari, ma nella lista delle richieste pubblicata sul sito, i gestori delle varie*

strutture postano desideri di tutti i gusti e per tutte le tipologie di clienti, in cambio di un soggiorno gratuito o perlomeno scontato. ”

Dopo: *“Questo è ciò che propone B&B di Quartu Sant' Elena in provincia di Cagliari , ma nella lista delle richieste pubblicata sul sito , i gestori delle varie strutture postano desideri di tutti i gusti e per tutte le tipologie di clienti , perlomeno scontato . ”*

In questo caso la frase dopo la semplificazione non ha mantenuto la sua grammaticalità, la manca un articolo “*un B&B*” è sufficiente a far perdere grammaticalità all’intera frase, in più la mancanza di “*in cambio di un soggiorno gratuito o perlomeno*” rende la fine della frase senza senso. La frase quindi dopo la semplificazione è stata considerata non grammaticale.

- **Prima:** *“ Incurante del nostro bisogno di adagiarci sui dogmi, la giornalista scientifica Michele Collet di Environmental Graffiti ha sfatato i luoghi comuni sul corpo umano, spiegando che non usiamo solo il 10% del cervello e che toccare un rospo non fa venire le verruche. ”*

Dopo: *“Incurante del nostro bisogno di adagiarci sui dogmi , la giornalista scientifica Michele Collet di Environmental Graffiti ha sfatato i luoghi comuni sul corpo umano , spiegando che non usiamo e che toccare un rospo non fa venire le verruche . ”*

Questo è un esempio concreto del fatto che abbiamo considerato frasi scorrette grammaticalmente in quanto dopo la semplificazione mancano parti del discorso che rendono la frase sensata. La mancanza di “*solo il 10% del cervello*” rende la frase priva di senso.

- **Prima:** “*Ma visto l'andazzo dell'economia e, perché no, anche la riscoperta di valori meno consumistici e il divertimento intrinseco nel provare una nuova esperienza, molti si sono appassionati e hanno deciso di prolungare l'avventura tutto l'anno.*”

Dopo: “*Ma visto l' andazzo dell' economia e , perché no , riscoperta di valori e il divertimento intrinseco nel provare una nuova esperienza , molti si sono appassionati e hanno deciso di prolungare l' avventura tutto l' anno.*”

In questo caso la frase dopo la semplificazione non ha mantenuto la grammaticalità, dopo la semplificazione manca “*anche la*”, e per questo la frase non risulta più grammaticale; la mancanza del modificatore “*meno consumistici*”, in isolamento non pregiudica la grammaticalità della frase.

Alcuni casi dubbi sono stati annotati con un tag specifico. Si tratta di frasi dove la mancanza di grammaticalità è data da un uso più marginale di alcune congiunzioni o da un segno di punteggiatura assente come ad esempio in questo caso:

“*Secondo la tradizione fu accecata e è considerata e festeggiata come la santa della luce.*”

In questo caso sarebbe preferibile usare la “*d*” eufonica: “*ed è*”, ma visto che per la semplificazione abbiamo usato solo il DELETE e la frase originale era:

“*Secondo la tradizione fu accecata e proprio per questo è considerata e festeggiata come la santa della luce.*”

Considerando che il DELETE non aggiunge parti del discorso, ma le cancella, abbiamo deciso di classificare questi tipi di errore come *forse*.

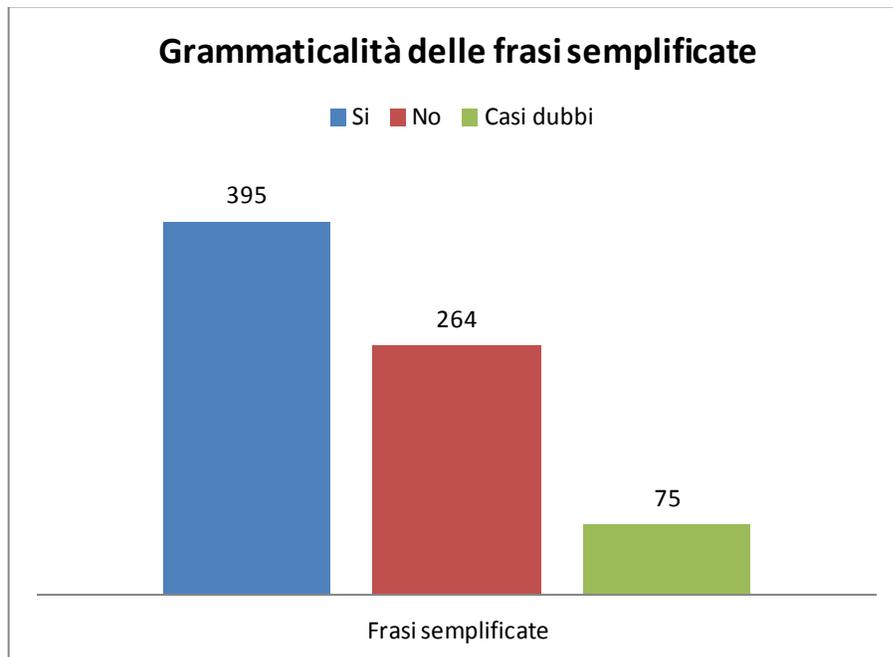


Figura 32: Grafico che riporta il numero di frasi grammaticali, non grammaticali e i casi dubbi.

La **figura 32** ci mostra i risultati dell'analisi sulla grammaticalità delle frasi semplificate, come possiamo notare dal grafico la maggioranza delle frasi continuano ad essere grammaticalmente corrette.

4.2.2 Analisi dell'informatività

La verifica dell'informatività è anch'essa avvenuta manualmente; le frasi semplificate sono state analizzate una alla volta, sia quelle grammaticali che quelle non grammaticali, verificando il mantenimento o meno dell'informatività del testo. Per aiutarci in questa valutazione abbiamo utilizzato le domande di cui abbiamo parlato nel capitolo 3.

Ecco alcuni esempi di frasi che hanno mantenuto o meno l'informatività:

- **Domanda:** “ Perché i bambini si nascondono? ”

Prima: “ *I bambini, preoccupati, si nascondono perché, secondo la leggenda, non possono vedere la Santa.* ”

Dopo: “ *I bambini , preoccupati , si nascondono perché non possono vedere la Santa .* ”

In questo caso la frase non solo ha mantenuto l'informatività, ma ha mantenuto anche la grammaticalità. Leggendo il testo con all'interno questa frase semplificata riusciamo a rispondere senza problemi alla domanda.

- **Domanda:** “ Il Fondo Mondiale per la Natura (WWF) è:
 - un'organizzazione internazionale
 - un parco nazionale del Tibet
 - una associazione che protegge i panda
 - il simbolo del panda ”

Prima: “ *Un po' per la sua rarità, un po' per il suo aspetto grazioso, quasi da animale di pezza, il panda è stato scelto come simbolo della più grande organizzazione internazionale per la conservazione degli animali, il Fondo Mondiale per la Natura (WWF).* ”

Dopo: “ *Un pò per la sua rarità , un pò per il suo aspetto grazioso , quasi da animale di pezza, il panda è stato scelto come simbolo della più grande*

organizzazione per la conservazione degli animali , il Fondo Mondiale per la Natura (WWF) . ”

In questo caso la frase mantiene la sua grammaticalità, ma perde informatività; infatti non è possibile dare una risposta alla domanda, anche se rimane “mondiale” che rende l’inferenza abbastanza semplice.

- **Domanda:** “ Chi capita dove vive il delfino? ”

Prima: “ *Un giorno, nella grotta dove nuotava il delfino capitò un sommozzatore, che cercava pesci-luna; la grotta ne era piena.* ”

Dopo: “ *Un giorno, la grotta ne era piena.* ”

In questo caso la frase semplificata è grammaticalmente corretta, ma perde la sua informatività in modo evidente, infatti non possiamo rispondere alla domanda.

- **Domanda:** “ Qual è la prima cosa che il cliente deve osservare con attenzione quando entra in un bar? ”

- deve osservare che tutto sia pulito e ordinato
- deve osservare il barista
- deve guardare la macchina del caffè ”

Prima: “ *Quando entriamo in bar è importante fare attenzione a ciò che vediamo: tutto deve essere pulito e ordinato.* ”

Dopo: “ *Quando entriamo in bar è importante : tutto deve essere pulito e ordinato.* ”

In questo caso la frase analizzata è un po' anomala, la mancanza di “*fare attenzione a ciò che vediamo*” rende la frase senza senso; tuttavia riusciamo a rispondere alla domanda in quanto la comprensione del testo è ancora possibile.

- **Domanda:** “Una delle malattie che lupi e cani possono trasmettere è
 - gli orecchioni
 - i pidocchi
 - l’influenza intestinale
 - la rabbia”

Prima: “*Questi ultimi soprattutto nel Centro Sud, sono diventati ormai un pericolo per gli animali domestici e selvatici, oltre che un veicolo di diffusione di malattie parassitarie e della rabbia.*”

Dopo: “*Questi ultimi soprattutto nel Centro Sud , sono diventati ormai un pericolo , oltre che un veicolo di diffusione di malattie parassitarie e della rabbia.*”

In questo caso la frase ha mantenuto la sua grammaticalità ed inoltre risponde alla domanda di comprensione del testo, tuttavia possiamo dire che c'è una perdita di informatività dovuta alla mancanza di “*per gli animali domestici e selvatici*”; questa parte della frase, anche se non contemplata nelle domande, potrebbe essere ritenuta importante per la comprensione del testo.

L'analisi dell'informatività non si è basata solamente sul fatto che una frase rispondesse o meno ad una delle domande relative al testo, questo perché gran parte delle frasi non avevano niente a che fare con le domande poste per la comprensione del testo.

Come abbiamo notato nell'esempio precedente ci sono anche frasi che nonostante rispondano in modo esaustivo alle domande di comprensione, perdono comunque una parte della loro informatività; proprio per questo le frasi sono state analizzate sotto due punti di vista:

1. Risponde alle domande?
2. La parte del testo eliminata contiene una quantità d'informazione o una tipologia di informazioni tale da far perdere informatività al testo?

La prima domanda è mirata sulla singola frase presa separatamente rispetto al resto del testo, la seconda domanda invece è riferita alla frase all'interno del testo e quindi al mantenimento dell'informatività del testo in generale. Per questo abbiamo risposto a queste due domande separatamente e abbiamo quindi creato su Excel due colonne separate in cui sono riportati i risultati.

Per quanto riguarda la prima domanda, come per la grammaticalità, anche in questo caso abbiamo risposto a queste domande con “*si*” se la frase risponde alle domande, con “*no*” in caso contrario e con “*forse*” nel caso in cui ci fosse una risposta parziale alla domanda.

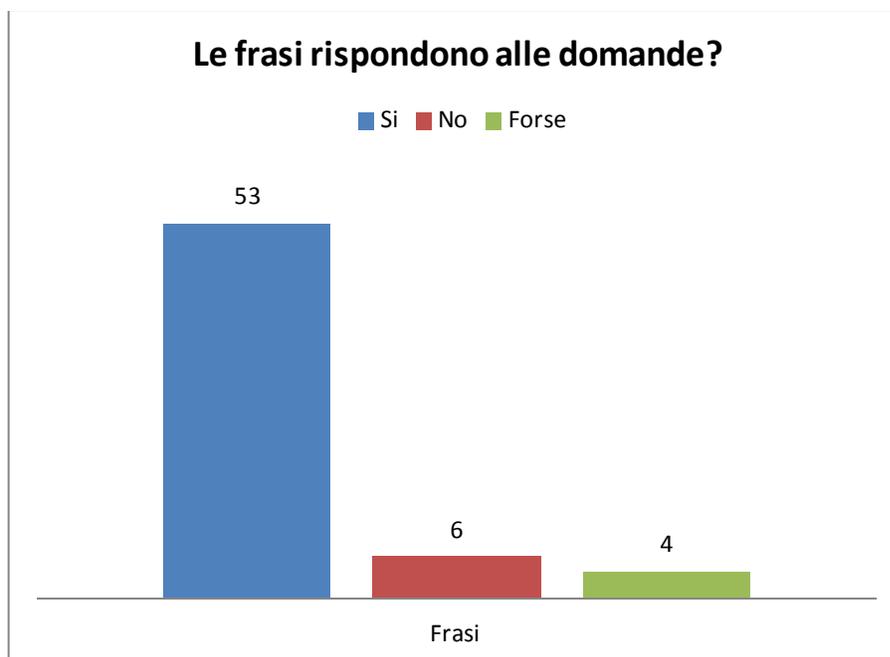


Figura 32: Grafico che rappresenta il numero di frasi che rispondono o meno alle domande di comprensione del testo.

Come possiamo notare dalla **figura 32**, solo poche delle frasi semplificate rispondevano direttamente ad una delle domande che abbiamo trovato per la comprensione del testo, ma di queste la maggior parte ha dato un risultato positivo.

Proprio perché la quantità di domande relative alle frasi prese in analisi erano poche abbiamo deciso di fare un secondo tipo di analisi sull'informatività delle frasi del testo semplificato, parliamo appunto della seconda domanda.

Quindi per quanto riguarda la seconda domanda, abbiamo risposto con “*ok*” nel caso in cui la parte di frase eliminata non contenesse informazioni utili alla comprensione del testo in generale e quindi la frase mantenesse la propria informatività, con “*no*” nel caso in cui la parte di testo eliminata contenesse informazioni da noi ritenute rilevanti per la comprensione del testo, e con “*forse*” nel caso in cui la parte di testo eliminata fosse solo una parola o un aggettivo che potrebbe risultare utile per la comprensione del testo nel complesso.

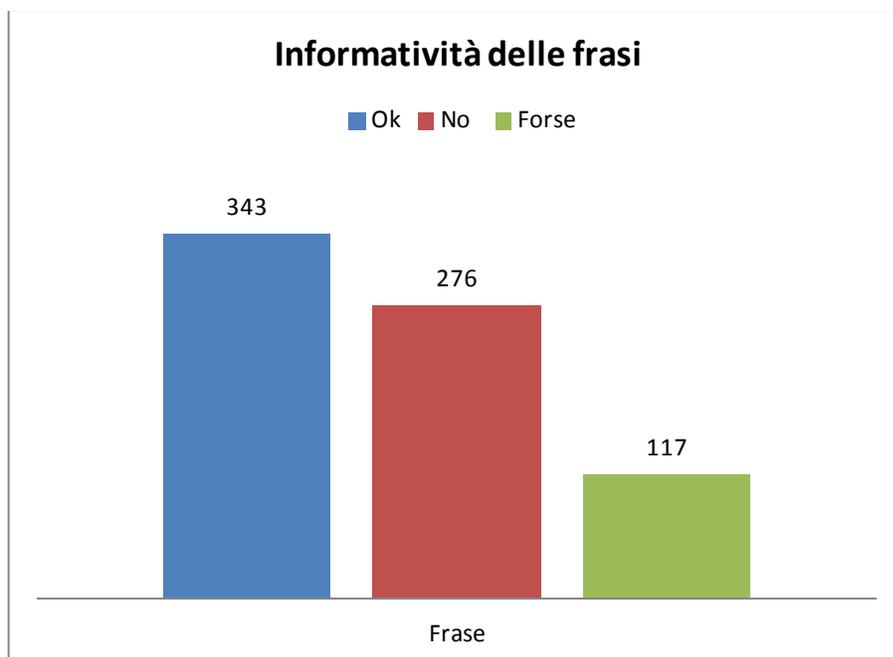


Figura 33: Grafico che rappresenta il numero di frasi che hanno mantenuto l'informatività dopo la semplificazione.

Come possiamo osservare in **figura 33**, sono state analizzate tutte le frasi semplificate per determinare il mantenimento o meno dell'informatività; possiamo notare che anche in questo caso la maggior parte delle frasi semplificate hanno mantenuto la loro informatività rispetto alla comprensione del testo.

5. CONCLUSIONI

Questo lavoro si è posto come obiettivo la definizione di un metodo e la creazione di un corpus per la valutazione della semplificazione automatica del testo che, nel nostro caso abbiamo ottenuto con l'applicazione della regola di cancellazione (delete); tuttavia il corpus che abbiamo creato e i metodi di valutazione che abbiamo applicato sono generali, possono cioè essere utilizzati per testare e valutare qualsiasi regola di semplificazione.

I metodi che abbiamo utilizzato per la valutazione del testo semplificato si possono dividere in automatici e manuali: i metodi automatici, nel nostro caso, sono quelli che riguardano l'utilizzo di un o strumento chiamato READ-IT per la valutazione della complessità di un testo; questo strumento è stato utilizzato sulle frasi prima della semplificazione e dopo la semplificazione per poter così confrontare l'effetto della semplificazione stessa. I metodi manuali sono la valutazione della grammaticalità di un testo, avvenuta esaminando frase per frase ogni testo semplificato, e la valutazione del mantenimento dell'informatività del testo dopo la semplificazione, avvenuto in parte grazie alle domande associate al testo ed in parte con una valutazione della parte di frase eliminata.

Il corpus che abbiamo creato è formato da testi per studenti italiani di vario ordine e grado che sono stati selezionati tra quelli delle prove INVALSI e tra testi di verifica dell'apprendimento; in particolare i testi che abbiamo preso in considerazione sono quelli a cui sono associate domande sulla comprensione del testo. Nella creazione del nostro corpus abbiamo diviso i testi in cartelle a seconda della classe scolastica a cui appartengo e contemporaneamente abbiamo diviso i testi dalle domande, assicurandoci che fossero domande relative alla comprensione del testo perché ci sono serviti per la valutazione dell'informatività dei testi semplificati.

I testi appartenenti al nostro corpus sono 137 e sono stati analizzati inizialmente prima della semplificazione tramite uno strumento per la valutazione automatica della leggibilità chiamato READ-IT, che ci ha fornito i diversi valori di complessità dei testi in generale e di ogni singola frase. Grazie a questa prima analisi abbiamo potuto osservare che nella maggior parte dei casi, la difficoltà dei testi analizzati aumenta con l'aumentare della classe scolastica.

Dopo questa prima analisi dei testi abbiamo applicato la nostra regola di semplificazione automatica del testo, il delete, ed abbiamo analizzato i risultati a diversi livelli: per quanto riguarda l'analisi della complessità avvenuta con READ-IT, abbiamo confrontato i risultati dei testi di partenza con quelli ottenuti dopo la semplificazione automatica dei testi; abbiamo notato che la semplificazione del testo con il delete ha influito, anche se di poco, sui diversi livelli di complessità analizzati, infatti tra i testi di partenza ed i testi semplificati vi è una differenza quasi impercettibile.

Per quanto riguarda i risultati dell'analisi della grammaticalità delle frasi appartenenti ai testi semplificati con l'applicazione del delete, la maggior parte delle frasi semplificate risultano grammaticalmente corrette. Infine, per quanto riguarda l'informatività delle frasi semplificate abbiamo diviso l'analisi in due parti: la prima analisi è stata fatta per verificare che le frasi rispondessero alle domande di comprensione associate al testo, la seconda è stata fatta per verificare se la parte della frase eliminata contenesse o meno informazioni importanti per la comprensione del testo; questo perché solo una parte delle frasi prese in analisi erano riferite alle domande di comprensione del testo.

Analizzando i risultati sull'informatività delle frasi che contenevano le risposte alle domande di comprensione del testo, possiamo dire che la maggior parte delle frasi semplificate con il delete hanno mantenuto l'informatività. Per quanto riguarda l'analisi dell'informatività delle frasi in generale quindi senza tener conto delle domande, abbiamo ottenuto risultati positivi, cioè anche in questo caso la maggior parte delle frasi ha mantenuto la sua informatività.

In conclusione analizzando nel complesso i nostri risultati, possiamo dire che la valutazione della regola di semplificazione “delete” ha dato risultati positivi: i testi anche se di poco hanno diminuito il loro grado di complessità, la maggior parte delle frasi semplificate hanno mantenuto la loro grammaticalità e la loro informatività.

Bibliografia

- Biljana Drndarević, Stanja Štajner, Stefan Bott, Susana Bautista, and Horacio Saggion. 2013. *Automatic text Simplification in Spanish: A Comparative Evaluation of Complementing Modules*. A. Gelbukh (ed.): CICLing 2013, Part II, LNCS 7817, pp. 488-500.
- Dominique Brunato, Felice Dell’Orletta, Giulia Venturi, Simonetta Montemagni. 2015. *Design and Annotation of the First Italian Corpus for text Simplification*. Proceedings of LAW IX - The 9th Linguistic Annotation Workshop, pp. 31–41. Denver, Colorado.
- Felice Dell’Orletta, Simonetta Montemagni and Giulia Venturi. 2011. *READ-IT: assessing Readability of Italina Text with a View to Text Simplification*. Proceedings of the 2nd Workshop on Speech and Language Processing for Assistive Technologies, pp. 73-83.
- Horacio Saggion, Sanja Štanjer, Stefan Bott, Simon Mille, Luz Rello, Biljana Dindarevic. 2015. *Making it Simplex: implementation and Evaluation of a text Simplification System for Spanish*. ACM Transactions on Accessible Computing, vol6.
- Irina Temnikova, Galina Maneva. 2013. *The C-Score –Proposing a Reading Comprehension Metrics ad a Common Evaluation Measure for text Simplification*. Proceedings of the 2nd Workshop on Predicting and Improving Text Readability for Target Reader Populations , pp.20-29.
- Sowmya Vajjala and Detmar Meurers. *Assessing the relative reading level of sentence pairs for text simplification*.
- Stefan Bott, Horacio Saggion. 2014. *Text simplification resources for Spanish*. Springer Science+Business Media Dordrecht.
- Walter S. Lasecki, Luz Rello, Jeffrey P. Bigham. *Measuring Text Simplification with the Crowd*. ACM.

Siti Web

- Alphacentauri, testi inseriti nella cartella Vario
<http://www.alphacentauri.it> (visitato il 24 giugno 2017)
- Grammatica Italiana, testi inseriti nell'omonima cartella
<http://grammaticaitaliana.net> (visitato il 24 giugno 2017)
- Invalsi, Istituto nazionale per la valutazione del sistema educativo di istruzione e di formazione
<http://www.invalsi.it/invalsi/index.php> (visitato il 24 giugno 2017)
- Italiano per stranieri, testi inseriti nella cartella Stranieri
<http://italianoperstranieri.loescher.it> (visitato il 24 giugno 2017)
- Maestra di tutto un po, testi inseriti nella cartella Vario
<http://maestradituttounpo.blogspot.it> *<http://www.alphacentauri.it>* (visitato il 24 giugno 2017)
- Maestra Sabry, testi inseriti nella cartella ItalianoCB
<http://www.maestrasabry.it> (visitato il 24 giugno 2017)
- READ-IT, strumento di valutazione della leggibilità
<http://www.italianlp.it/demo/read-it/> (visitato il 24 giugno 2017)
- Risorse didattiche, testi inseriti nell'omonima cartella
<http://www.risorsededidattiche.net> (visitato il 24 giugno 2017)
- Schola, testi inseriti nella cartella
<http://share.dschola.it> (visitato il 24 giugno 2017)
- Scuole Valnervia, testi inseriti nella cartella Prove
<http://www.scuolevalnervia.org> (visitato il 24 giugno 2017)
- Servimedia, agenzia spagnola
<http://www.servimedia.es/> (visitato il 24 giugno 2017)

Ringraziamenti

Ringrazio i miei relatori la dott.ssa Dominique Brunato e il dott. Felice Dell'Orletta che sono stati sempre disponibili ad aiutarmi e ad indirizzarmi rendendo così possibile la realizzazione di questa tesi.

Ringrazio mia madre per avermi permesso di continuare gli studi universitari grazie ai suoi sacrifici, ed avermi sostenuta anche quando ho deciso di cambiare facoltà per poter trovare la mia strada. Sono stati anni difficili pieni di alti e bassi, ma nonostante tutto mi ha sempre aiutata ed incoraggiata a fare di più.

Ringrazio mia zia Mariella e mia nonna che si sono sempre interessate ai miei studi, mi hanno sempre sostenuta ed incoraggiata e sono sempre state orgogliose di tutti i miei progressi.

Ringrazio tutti i miei parenti per avermi sostenuto ed incoraggiato a finire gli studi, nonostante gli alti e bassi che ho avuto in questi anni.

Ringrazio Fabricio per avermi incoraggiata in questo ultimo anno e per essermi stato vicino anche nei momenti bui, per tutti i sorrisi e le risate, per le nostre cene e i nostri pranzi per i nostri viaggi brevi e lunghi, e soprattutto per sopportarmi ed essere sempre al mio fianco.

Ringrazio gli amici conosciuti a Pisa Davide, Ivan, Gianmaria, Manuele e Roberto che mi sono stati accanto in ogni momento bello o brutto durante tutti questi anni e mi hanno incoraggiata sempre a fare del mio meglio anche nei momenti difficili e sono sempre riusciti a strapparmi un sorriso ed una risata. Ricorderò sempre le nostre cene del lunedì sera, tutte le nostre serate folli.

Ringrazio gli amici di Spezia per le risate e la spensieratezza delle nostre serate insieme, per avermi sempre strappato un sorriso ed incoraggiata a fare del mio meglio.