


Politecnico di Torino


## Politecnico di Torino

Corso di Laurea Magistrale in Ingegneria Informatica

A.a. 2025/2026

Sessione di laurea Ottobre 2026


# ReadLens: un'applicazione web per la lettura e l'apprendimento dell'inglese assistiti da IA


Relatori:

Luigi De Russis

Tommaso Calò

Candidato:

Giulio Guarino

# Ringraziamenti

Ringrazio sentitamente i miei relatori Luigi De Russis e Tommaso Calò per i preziosi consigli e il sostegno che mi hanno dato, e per la loro disponibilità nell'accompagnarmi in questo percorso di tesi.

Ringrazio tutti i partecipanti allo studio formativo e allo studio conclusivo del progetto, e le professoresse del CLA e di UniTo che hanno deciso di partecipare allo studio iniziale dedicandomi parte del loro tempo.

Infine, ringrazio di cuore la mia famiglia e i miei amici. Senza di voi niente di tutto ciò sarebbe possibile.

# Indice

**Elenco delle tabelle** VII

**Elenco delle figure** VIII

**Glossario** IX

**1 Introduzione** 1
- 1.1 Obiettivi . . . 2
- 1.2 Struttura della tesi . . . 3

**2 Background** 5
- 2.1 Lettura e acquisizione del lessico in L2 . . . 5
- 2.2 L'Involvement Load Hypothesis . . . 6
- 2.3 La lettura come processo: fluenza, motivazione, lettura estensiva . . . 7
- 2.4 Strumenti digitali di supporto alla lettura in L2 . . . 9
  - 2.4.1 Il glossing come supporto alla lettura in L2 . . . 9
  - 2.4.2 Dizionari e traduttori integrati . . . 10
  - 2.4.3 Strumenti per la lettura in L2 basati su NLP e LLM . . . 11

**3 Studio Formativo** 12
- 3.1 Domande di ricerca . . . 12
- 3.2 Metodo . . . 13
  - 3.2.1 Partecipanti . . . 13
  - 3.2.2 Strumenti di raccolta dei dati . . . 13
  - 3.2.3 Approccio di analisi . . . 14
- 3.3 Profilo del campione . . . 15
  - 3.3.1 Sondaggio (n=89) . . . 15
  - 3.3.2 Interviste con lettori . . . 15
- 3.4 Risultati . . . 16
  - 3.4.1 Difficoltà nella lettura in inglese . . . 16
  - 3.4.2 La lettura come occasione di apprendimento . . . 19

3.4.3 Bisogni durante la lettura in inglese . . . 21
3.4.4 Come la lettura può supportare l'acquisizione linguistica . . 24
3.4.5 Sintesi trasversale . . . 26
3.5 Limiti dello studio formativo . . . 27

**4 Definizione dei Requisiti** 29
4.1 Personas e Context Scenarios . . . 29
4.1.1 Definizione delle personas . . . 29
4.1.2 Definizione dei context scenarios . . . 31
4.2 Sintesi dei bisogni e definizione dello scope . . . 33
4.2.1 Bisogni emersi . . . 33
4.2.2 Definizione delle priorità e scelte di scope . . . 35
4.3 Formalizzazione dei requisiti . . . 36
4.3.1 Requisiti funzionali . . . 36
4.3.2 Requisiti non funzionali . . . 37

**5 Progettazione** 39
5.1 Decisioni di design . . . 39
5.1.1 Approccio inline . . . 40
5.1.2 Lingua nello strumento . . . 40
5.1.3 Livelli di approfondimento . . . 41
5.1.4 Classificazione automatica . . . 41
5.1.5 Persistenza delle interazioni . . . 42
5.2 Wireframes . . . 42
5.2.1 Interazione minima . . . 43
5.2.2 Pannello inline . . . 45
5.2.3 Selezione di un prompt . . . 46
5.2.4 Classificazione automatica e manuale . . . 47
5.2.5 Modalità Esercitati . . . 49
5.2.6 Note persistenti . . . 50
5.3 Architettura del sistema . . . 50

**6 Implementazione** 52
6.1 Scelte progettuali in corso d'opera . . . 52
6.2 Tecnologie . . . 53
6.3 Frontend . . . 54
6.3.1 Architettura e organizzazione . . . 54
6.3.2 Linee guida per l'interazione con l'IA . . . 55
6.3.3 Stati dell'interfaccia . . . 56
6.3.4 Componenti principali . . . 66
6.4 Backend . . . 67

6.4.1 Architettura e organizzazione . . . . . 67
6.4.2 Generazione dei contenuti . . . . . 68
6.4.3 Gestione delle conversazioni . . . . . 69
6.4.4 Persistenza dei dati . . . . . 70

**7 Valutazione** 72
7.1 Metodo . . . . . 72
7.1.1 Domande di ricerca . . . . . 72
7.1.2 Partecipanti . . . . . 73
7.1.3 Materiali . . . . . 73
7.1.4 Strumenti di misura . . . . . 74
7.1.5 Procedura . . . . . 77
7.1.6 Analisi dei dati . . . . . 79
7.2 Risultati . . . . . 80
7.2.1 Risultati della comprensione (RQ1) . . . . . 81
7.2.2 Risultati dell'apprendimento (RQ2-RQ5) . . . . . 81
7.2.3 Risultati di engagement e costo percepito (RQ3-RQ4) . . . . 84
7.2.4 Risultati interviste . . . . . 87
7.3 Discussione dei risultati . . . . . 90
7.4 Limiti dello studio . . . . . 91

**8 Conclusioni** 94
8.1 Limitazioni e lavoro futuro . . . . . 95

**Bibliografia** 97

# Elenco delle tabelle


3.1 Distribuzione del campione per livello CEFR . . . . . 15
3.2 Difficoltà nella lettura per livello CEFR . . . . . 16
3.3 Bisogni durante la lettura in inglese . . . . . 22

4.1 Bisogni emersi dall'analisi dei context scenarios, condivisi e per persona. . . . . 34
4.2 Requisiti funzionali del sistema. . . . . 37
4.3 Requisiti non funzionali del sistema. . . . . 38

6.1 Linee guida di Amershi et al. (2019) richiamate nel capitolo. . . . . 55

7.1 Sequenza delle attività nella seconda sessione dello studio. . . . . 78
7.2 Sintesi dei risultati statistici per domanda di ricerca. Le righe in grassetto indicano risultati statisticamente significativi ($\alpha = .05$). . 87

# Elenco delle figure


5.1 Stato iniziale dell'interfaccia di lettura. . . . . 43
5.2 Selezione di una parola e risposta immediata. . . . . 44
5.3 Pannello inline espanso. . . . . 45
5.4 Prompt selezionato . . . . . 46
5.5 Navigazione tra risposte ricevute . . . . . 47
5.6 Risposta salvata nella nota . . . . . 47
5.7 Apertura menù classificazione . . . . . 48
5.8 Categoria possibilmente sbagliata . . . . . 48
5.9 Modalità "Esercitati" . . . . . 49
5.10 Nota inline . . . . . 50
5.11 Architettura concettuale del sistema . . . . . 51

6.1 Libreria con alcuni testi già importati. . . . . 56
6.2 Form per l'aggiunta di un nuovo testo. . . . . 57
6.3 Schermata di lettura . . . . . 58
6.4 Interazioni del primo livello . . . . . 59
6.5 Pannello inline appena aperto . . . . . 60
6.6 Apertura input per domanda libera . . . . . 60
6.7 Generazione risposte in seguito a click su prompt di default . . . . . 61
6.8 Seconda interazione tramite domanda di followup . . . . . 62
6.9 Possibilità di lasciare un feedback . . . . . 63
6.10 Menu a tendina per correggere la classificazione . . . . . 64
6.11 Avviso di categoria potenzialmente sbagliata . . . . . 64
6.12 Il testo con cui l'utente interagisce rimane evidenziato . . . . . 65
6.13 La nota si apre a lato del testo . . . . . 65
6.14 Apertura di risposte ricevute in passato nella nota . . . . . 66
6.15 Diagramma entità-relazione del modello dati. . . . . 70

# Glossario

**LLM**

Large Language Model

**CALL**

Computer Assisted Language Learning

**NLP**

Natural Language Processing

**SLA**

Second Language Acquisition

**CLA**

Centro Linguistico d'Ateneo

**CEFR**

Common European Framework of Reference

**EFL**

English as a Foreign Language

**ILH**

Involvement Load Hypothesis

**L1**

La lingua acquisita per prima dal parlante, tipicamente in un contesto naturale durante l'infanzia (spesso, ma non necessariamente, coincidente con la cosiddetta "lingua madre").

**L2**

Una lingua appresa dal parlante successivamente alla L1, spesso in un contesto istruzionale; nella letteratura di settore il termine è usato come categoria generale per indicare qualunque lingua non nativa, non necessariamente la seconda in senso stretto cronologico.

**graded reader**

testo semplificato con vocabolario controllato, calibrato su un livello di competenza

**apprendente**

Individuo impegnato nel processo di acquisizione di una lingua, sia essa la L1 che una L2, in un contesto formale (istruzionale) o informale.

**lingua target**

La lingua straniera (L2) che l'apprendente sta attivamente cercando di apprendere, distinta da eventuali altre lingue coinvolte nel contesto di apprendimento o insegnamento.

**wireframe**

Rappresentazione a bassa fedeltà della struttura e del comportamento di un'interfaccia, priva di dettagli grafici definitivi.

# Capitolo 1

# Introduzione

L'interesse verso il potenziale dell'intelligenza artificiale nell'assistere gli utenti nelle loro attività quotidiane, personalizzandone l'esperienza e restando sempre disponibile, ha ricevuto una spinta decisiva negli ultimi anni con il rilascio pubblico di ChatGPT, a cui ha fatto seguito una rapida diffusione di LLM generativi ormai entrati nella nostra quotidianità. Parallelamente, si è assistito a una rapida crescita della ricerca HCI sul loro impiego in sistemi e applicazioni rivolti agli utenti, come evidenziato dal marcato aumento di lavori dedicati agli LLM nella comunità HCI [1].

Una svolta analoga si è verificata nell'ambito dell'apprendimento delle lingue. Apprendere una lingua straniera resta, per la maggior parte delle persone, un processo lento e non banale, specialmente in assenza di occasioni reali di immersione nella lingua target o di supporto specialistico. È proprio su questo fronte che l'IA generativa ha il potenziale di intervenire, abbassando il costo di accesso a un supporto che ne faccia le veci. Dal 2022 in poi, non a caso, la ricerca sull'apprendimento linguistico mediato da tecnologie IA (una naturale estensione del più ampio campo del Computer-Assisted Language Learning, CALL) è cresciuta in modo marcato: una recente rassegna sistematica documenta una crescita quasi esponenziale degli studi pubblicati nel biennio successivo al 2022, che riportano benefici nell'apprendimento delle lingue che attraversano dimensioni linguistiche, affettive e cognitive, inclusa una riduzione dell'ansia comunicativa dovuta in parte all'assenza del giudizio che l'interazione con un interlocutore umano può comportare [2].

Tra le attività il cui beneficio per l'apprendimento linguistico è meglio documentato c'è la lettura. Questa espone il lettore a una grande varietà di forme lessicali e costruzioni grammaticali inserite in contesti autentici, offrendo un tipo di esposizione alla lingua che lo studio in contesti formali, quale può essere quello in aula con insegnante, da solo non può fornire nella stessa misura.

Non tutti gli apprendenti riescono però a sfruttare questa risorsa. Leggere in

una seconda lingua (L2) richiede un impegno maggiore rispetto alla lettura nella prima lingua (L1), non solo per la barriera linguistica in sé ma perché ogni parola sconosciuta costringe il lettore a una scelta: consultare una risorsa esterna (un dizionario, un traduttore, un LLM), interrompendo il flusso di lettura, oppure tralasciare l'incomprensione, con il rischio di non cogliere l'interezza del testo e di ridurre il piacere della lettura, e con essa la probabilità che questa venga svolta con consistenza. Il problema dell'interruzione non è esclusivo della lettura in L2, ma qui ha conseguenze marcate. La scorrevolezza della lettura infatti si costruisce proprio sulla capacità di procedere senza fermarsi a ragionare su ogni singola parola, e ogni interruzione per cercarne il significato spezza quella continuità. Questo costo, inoltre, pesa di più quando le risorse cognitive del lettore sono già impegnate nel tentativo di comprendere e assorbire una lingua che non è la sua.

Gli strumenti oggi disponibili risolvono solo in parte questo attrito [3, 4]. Le glosse elettroniche, la forma di supporto lessicale alla lettura più consolidata in letteratura SLA (Second Language Acquisition), riducono l'interruzione rendendo il supporto immediatamente accessibile vicino alla parola, ma restano vincolate a un insieme di parole stabilito in anticipo e a una spiegazione scritta per un lettore generico, non per il dubbio specifico che un lettore incontra in un dato momento [5, 6]. Dizionari e traduttori integrati, come quelli offerti da lettori digitali quali il Kindle, superano questo primo limite permettendo di cercare qualunque termine al momento del bisogno, ma restano decontestualizzati e propongono significati generici o traduzioni isolate, spesso poco pertinenti all'incomprensione specifica riscontrata dal lettore in quella frase, lasciando a lui il compito di individuare quale accezione del significato si applichi. Ricorrere a un LLM come ChatGPT risolve meglio il problema della contestualizzazione, ma tipicamente al prezzo di uscire dall'esperienza di lettura. L'utente deve infatti aprire uno strumento separato, formulare correttamente una richiesta, e spesso ripetere o integrare manualmente il contesto per ottenere una spiegazione utile, reintroducendo proprio l'interruzione che il supporto dovrebbe ridurre [4].

## 1.1 Obiettivi

Il progetto esplora come l'intelligenza artificiale generativa possa essere integrata nell'interazione con un testo in L2 per offrire supporto alla comprensione e all'apprendimento, senza separare l'utente dal contesto in cui incontra un contenuto incompreso. In particolare intende realizzare un supporto alla lettura rivolto a lettori italiani di inglese: una scelta motivata sia dalla rilevanza pratica, poiché l'inglese è presumibilmente la lingua straniera più richiesta in ambito universitario italiano, sia dalla maggiore reperibilità, nello stesso ambiente universitario, di potenziali utenti target.

La letteratura sulla relazione tra lettura e acquisizione lessicale in L2, pur offrendo un quadro teorico solido, non permette da sola di stabilire come i suoi principi possano tradursi in scelte di design dell'interazione, né chiarisce quali forme di supporto siano maggiormente adatte ai bisogni concreti dei lettori italiani di inglese. Di conseguenza, la tesi indaga innanzitutto, tramite uno studio formativo, i bisogni, le pratiche e gli ostacoli attuali di questi lettori, e ne usa i risultati per progettare, implementare e valutare uno strumento di supporto alla lettura in L2 che ne faciliti la comprensione e ne sostenga l'apprendimento.

Lo strumento realizzato è ReadLens, un'applicazione web basata su intelligenza artificiale che offre supporto lessicale e contestuale durante la lettura di testi in inglese, integrando la spiegazione di parole ed espressioni direttamente nel flusso della lettura, senza costringere il lettore a interromperla per consultare risorse esterne.

Il suo contributo alla comprensione e all'apprendimento del lessico, come anche il carico di lavoro e l'usabilità percepiti, sono stati verificati tramite uno studio controllato con partecipanti, confrontandoli con una soluzione di riferimento basata su strumenti generici già esistenti.

## 1.2 Struttura della tesi

Il processo di progettazione adottato si è ispirato all'approccio del Goal-Directed Design proposto da Cooper et al. in About Face [7], che articola il processo in una sequenza di fasi partendo dall'analisi dei bisogni degli utenti, proseguendo con la definizione dei requisiti e delle scelte di design, arrivando infine all'implementazione e valutazione del prodotto realizzato. Questa tesi riflette perciò l'ordine di questo approccio ed è strutturata nel modo seguente:

**Capitolo 2, Background:** analizza la letteratura sulla lettura come canale di acquisizione del lessico in L2, dalle condizioni che rendono efficace un incontro con una parola nuova fino agli strumenti digitali esistenti a supporto della lettura, discutendone le potenzialità e i limiti. Questo capitolo fornisce il quadro teorico su cui si fondano le scelte discusse nei capitoli successivi.

**Capitolo 3, Studio Formativo:** descrive lo studio condotto con lettori italiani di inglese ed esperte di insegnamento della lingua, volto a comprendere le difficoltà, i bisogni e le abitudini che caratterizzano la lettura in una lingua straniera. I risultati, discussi anche alla luce della letteratura presentata nel capitolo precedente, costituiscono la base empirica da cui muove la progettazione dello strumento.

**Capitolo 4, Definizione dei Requisiti:** traduce i risultati dello studio formativo in personas e scenari d'uso rappresentativi, da cui vengono derivati i requisiti funzionali e non funzionali dello strumento.

**Capitolo 5, Progettazione:** presenta le decisioni di design che traducono i requisiti in un'interfaccia concreta, illustrate attraverso dei wireframes, e l'architettura generale del sistema ideato.

**Capitolo 6, Implementazione:** descrive come il design è stato realizzato in un prototipo funzionante: le tecnologie impiegate, l'architettura di backend e frontend, l'organizzazione della persistenza dei dati, e il funzionamento dello strumento risultante.

**Capitolo 7, Valutazione:** riporta il protocollo, i risultati e la discussione dello studio condotto per valutare il contributo dello strumento rispetto a una soluzione di riferimento, sia in termini di carico di lavoro e usabilità percepiti sia di supporto effettivo alla comprensione e all'apprendimento del lessico.

**Capitolo 8, Conclusioni:** riassume i contributi del lavoro, ne discute i limiti e delinea le direzioni di sviluppo future.

# Capitolo 2

# Background

Questo capitolo presenta il quadro teorico in cui si posiziona il progetto: i meccanismi con cui il vocabolario si acquisisce durante la lettura in L2, le condizioni che ne determinano l'efficacia, e gli strumenti digitali esistenti per supportare questo processo.

## 2.1 Lettura e acquisizione del lessico in L2

Il rapporto tra vocabolario e comprensione della lettura è bidirezionale: un vocabolario più ampio rende accessibili testi altrimenti troppo difficili, e la lettura è a sua volta uno dei canali principali attraverso cui il vocabolario stesso si espande [8]. Questo legame, definito in letteratura *reciprocal causation*, rende la lettura un canale privilegiato di acquisizione lessicale sia nella lingua madre (L1) che in una seconda lingua (L2).

Il meccanismo attraverso cui il vocabolario viene acquisito durante la lettura è quello dell'*incidental learning*: le parole vengono apprese come effetto collaterale di un'attività il cui scopo primario è la comprensione del testo, non lo studio della lingua [9]. Si tratta di un processo lento e incrementale, in cui l'esposizione ripetuta a parole in contesti diversi permette di costruire gradualmente una conoscenza lessicale stabile. A questo si contrappone l'*intentional learning*, ovvero lo studio deliberato del vocabolario attraverso definizioni, flashcard o esercizi mirati. È una modalità più efficace per ottenere risultati rapidi ma non sufficiente da sola a sviluppare il repertorio lessicale esteso necessario per leggere testi autentici [10], per questo le due modalità sono generalmente considerate complementari.

In L2, tuttavia, l'efficacia dell'acquisizione incidentale è meno netta e gli studi non convergono in maniera chiara. Alcuni documentano tassi di ritenzione scoraggianti anche dopo esposizioni multiple [11], altri un'acquisizione più consistente, misurata però su dimensioni diverse dal solo significato [12]. In generale, molti apprendenti di

L2 non si trovano nelle condizioni necessarie perché questo processo possa avvenire in modo sistematico [6].Tra queste condizioni, un ruolo centrale spetta alla qualità del singolo incontro con la parola: è su questo aspetto che si concentra l'Involvement Load Hypothesis, discussa nella sezione seguente.

## 2.2 L'Involvement Load Hypothesis

Perché l'acquisizione incidentale abbia effettivamente luogo, è necessario che si verifichino alcune condizioni.

Una prima condizione riguarda la soglia di comprensione del testo: per mantenere un flusso di lettura funzionale, il lettore deve essere in grado di comprendere autonomamente una percentuale sufficiente delle parole presenti. Sono necessarie circa 4.000–5.000 famiglie di parole per coprire approssimativamente il 95% del vocabolario di un testo in inglese, una soglia al di sotto della quale la densità delle parole sconosciute compromette la comprensione complessiva, riducendo l'esposizione utile alla lingua prima che qualsiasi processo di acquisizione possa avviarsi [8].

Una seconda condizione riguarda la frequenza di esposizione. Un singolo incontro con una parola raramente è sufficiente per una ritenzione stabile: mentre il riconoscimento della forma può avvenire dopo poche occorrenze, una conoscenza solida del significato richiede incontri multipli, e il numero esatto varia in funzione della parola e del livello dell'apprendente, potendo oscillare tra 6 e oltre 20 occorrenze [13]. La varietà dei contesti è rilevante quanto la frequenza: rincontrare una parola in usi diversi produce un apprendimento più robusto rispetto alla ripetizione della stessa forma, perché favorisce il recupero attivo dalla memoria anziché la semplice rilettura della stessa informazione [14].

Una terza condizione riguarda la qualità del singolo incontro. Molti contesti risultano ambigui o poveri di informazioni. Per esempio uno studio ha rilevato che, su 70 parole sconosciute per le quali un contesto interpretabile era disponibile per 41, gli apprendenti ne hanno indovinate correttamente solo 17, portando [10] a concludere che, di norma, un apprendente riesce a indovinare correttamente meno della metà delle parole sconosciute incontrate in un testo. Non è dunque sufficiente esporre l'apprendente a testi contenenti parole nuove: è la qualità dell'incontro con quelle parole a determinare la probabilità di ritenzione.

Questo punto è al centro dell'Involvement Load Hypothesis (ILH), proposta da Laufer e Hulstijn [15] per spiegare perché alcune attività producano una ritenzione lessicale maggiore di altre. Il modello originale individua tre componenti che determinano il carico cognitivo (*involvement load*) associato a un incontro con una parola nuova:

- **Need**: la motivazione a comprendere una parola. È forte quando è autoimposta dall'apprendente, moderata quando è richiesta dal compito, assente negli altri casi.
- **Search**: il tentativo di reperire il significato di una parola sconosciuta attraverso una fonte adeguata (dizionario, persona competente o altro). È presente o assente.
- **Evaluation**: il confronto tra una parola e il suo contesto d'uso per valutarne il significato appropriato

Nella versione più recente del framework, denominata ILH Plus [16], la componente evaluation viene articolata in quattro livelli anziché tre, come previsti dall'originale, per distinguere gradazioni più fini di elaborazione produttiva: assenza di valutazione, quando il significato viene semplicemente fornito senza richiedere elaborazione da parte del lettore; valutazione semplice, quando il lettore confronta la parola con il proprio contesto d'uso senza produrne uno nuovo; uso variato a livello di frase, quando il lettore produce una singola frase che impiega la parola in un contesto diverso da quello originale; e uso variato a livello di composizione, il livello più elaborato, corrispondente all'integrazione della parola in un testo più esteso e coerente.

La meta-analisi conferma che i guadagni lessicali aumentano progressivamente lungo questi livelli: rispetto all'assenza di valutazione, una valutazione semplice produce un incremento contenuto, che cresce con l'uso variato a livello di frase e risulta massimo con l'uso variato a livello di composizione [16]. Anche il need contribuisce positivamente, mentre il search non risulta un predittore significativo dei guadagni immediati e mostra un effetto lievemente negativo, seppur non significativo, su quelli differiti.

Ne consegue che le attività a carico ridotto producono una ritenzione meno stabile nel tempo rispetto a quelle che richiedono un'elaborazione contestuale attiva del termine.

## 2.3 La lettura come processo: fluenza, motivazione, lettura estensiva

La lettura efficiente poggia su processi di basso livello ampiamente automatizzati, come il riconoscimento delle parole e l'analisi sintattica, che operano senza attenzione consapevole. L'efficienza della comprensione dipende dal fatto che questi processi non vengano interrotti da intrusioni consapevoli, poiché ogni deviazione dell'attenzione verso un'operazione deliberata ha un costo per il flusso complessivo

[8].Questa automaticità non è innata ma si sviluppa con la pratica, ed è attraverso l'esposizione a grandi quantità di testo che la fluenza di lettura si consolida.

È in questa prospettiva che si colloca la lettura estensiva (*extensive reading*), intesa come lettura di grandi quantità di testo il cui scopo primario resta la comprensione del contenuto. Si distingue dalla lettura intensiva (*intensive reading*), tipica delle attività d'aula guidate da un insegnante, in cui ogni porzione del testo viene analizzata nella sua forma; nell'extensive reading l'attenzione è sul significato, non sulla lingua come oggetto. È attraverso il volume che questa pratica dovrebbe soddisfare la condizione dell'incontro ripetuto discussa nella §2.2, esponendo il lettore a strutture linguistiche in uso e rinforzandole in modo incidentale.

Il contributo dell'extensive reading all'acquisizione è reale ma condizionato. A volumi sufficienti la lettura è un canale di acquisizione genuino: leggere circa un milione di parole in un anno può portare all'apprendimento di oltre 3.000 famiglie di parole nuove [8], e una meta-analisi documenta guadagni lessicali incidentali tra il 9 e il 18%, modesti ma consistenti [17]. Questi guadagni dipendono però strettamente dal volume e dalla qualità degli incontri. Nei contesti di apprendimento scolastico della L2 il volume di lettura è tipicamente molto più basso di quanto servirebbe: si stima che, se l'acquisizione dipendesse unicamente dall'esposizione tramite lettura in un contesto scolastico tipico, servirebbero circa 27 anni per accumulare 2000 parole [10]. Analogamente, Waring e Takaki [11] rilevano che dalla lettura di un singolo graded reader vengono appresi pochissimi termini nuovi, e circa metà di quelli appresi viene presto dimenticata, mentre il beneficio più stabile risiede nel consolidamento di parole già parzialmente note. Gli stessi autori sottolineano che questo non depone contro la lettura estensiva in sé, ma indica piuttosto la necessità di un volume di lettura molto più ampio di quello coperto da un singolo testo. La lettura si conferma quindi un canale di acquisizione autentico ma inefficiente, complementare all'attenzione deliberata al lessico più che sostitutivo, il cui rendimento dipende in misura decisiva dalla qualità del singolo incontro.

Perché la lettura estensiva produca benefici linguistici, il testo deve inoltre rispettare una soglia di copertura più esigente di quella minima discussa in §2.2: l'extensive reading richiede infatti che il lettore proceda in autonomia, senza il supporto lessicale continuo che caratterizza altri tipi di lettura assistita. La soglia proposta in letteratura è di circa una parola sconosciuta ogni 50, corrispondente al 98% di copertura, affinché la lettura possa restare un'esperienza sostenibile e non frustrante [18]. La dimensione motivazionale è quindi fondamentale poiché quando il carico lessicale supera la soglia di tolleranza, il lettore abbandona il testo o smette di elaborarlo attivamente, e l'esposizione si interrompe prima che qualsiasi apprendimento possa avvenire.

## 2.4 Strumenti digitali di supporto alla lettura in L2

Le sezioni seguenti presentano tre categorie di strumenti digitali per il supporto alla lettura in L2: il glossing elettronico, i dizionari e traduttori integrati, e gli strumenti più recenti basati su modelli linguistici di grandi dimensioni. Per ciascuna categoria vengono discussi il contributo specifico e i limiti residui rispetto al supporto alla lettura in L2.

### 2.4.1 Il glossing come supporto alla lettura in L2

Un gloss è una breve definizione o un sinonimo di una parola, in L1 o in L2, fornito insieme al testo [14]. Si tratta di una pratica antica (la glossa marginale risale al medioevo) che la ricerca CALL ha ripreso in forma elettronica. Nation [14] individua quattro funzioni del gloss nella lettura in L2. La prima è consentire l'uso di testi altrimenti troppo difficili per il lettore, rendendo accessibili testi autentici e non semplificati. La seconda è fornire il significato corretto di una parola, prevenendo le inferenze errate a cui il lettore andrebbe incontro affidandosi al solo contesto. La terza è ridurre al minimo l'interruzione del processo di lettura, specialmente quando la glossa compare vicino alla parola a cui si riferisce, un'efficienza che la consultazione di un dizionario tradizionale non garantisce. La quarta è richiamare l'attenzione sulla parola, favorendone potenzialmente l'apprendimento. A questo si aggiunge, nella tradizione elettronica del glossing, la possibilità per il lettore di scegliere quali glosse consultare tra quelle disponibili nel testo, rendendo il ritmo di consultazione personale.

La ricerca sul glossing elettronico è tra le più consolidate nel supporto lessicale alla lettura. Una meta-analisi su dieci studi riporta un effetto complessivo positivo e statisticamente significativo delle glosse ipertestuali sull'acquisizione del vocabolario. Il beneficio va però qualificato: i guadagni lessicali tendono a decadere già dopo due o tre settimane [19], coerentemente con la necessità di incontri ripetuti discussa in §2.3; inoltre non è la forma del gloss a determinarne l'efficacia, quanto il tipo di informazione fornita, con le glosse che comunicano direttamente il significato più efficaci di quelle che si limitano a stimolarne l'inferenza [20].

La lingua in cui la glossa è formulata è essa stessa una variabile, e la ricerca non converge su una scelta univoca. L'evidenza tende a favorire la L1 per la ritenzione: lo confermano sintesi qualitative [10] e studi sperimentali, che rilevano un vantaggio delle glosse L1 sia in comprensione sia in acquisizione del vocabolario [21]. In alcuni studi però questo vantaggio si dimostra moderato dalla competenza linguistica. Le glosse in L2 risultano più efficaci per i lettori di livello alto, mentre per i lettori di livello più basso la loro efficacia si riduce o viene superata da quella della L1 [22]. Le spiegazioni in L2, inoltre, offrono un'esposizione aggiuntiva alla lingua

e informazioni sul contesto d'uso che la traduzione non veicola [10]. La lingua ottimale della spiegazione non è quindi fissabile una volta per tutte: dipende dal livello del lettore e dall'obiettivo della lettura.

Resta, in ogni caso, un vincolo strutturale comune a tutta questa tradizione il fatto che le glosse sono tipicamente pre-redatte, con un insieme fisso di parole target e una spiegazione scritta in anticipo per un lettore generico, non per il dubbio specifico che un lettore incontra in un dato momento.

### 2.4.2 Dizionari e traduttori integrati

A differenza delle glosse pre-redatte, i dizionari, che siano cartacei, elettronici o online, svincolano il supporto dall'insieme di parole previsto in anticipo facendo sì che il lettore possa cercare qualunque termine nel momento in cui ne ha bisogno. I dizionari elettronici e online riducono lo sforzo di consultazione rispetto a quelli cartacei, e questo accesso più rapido tende ad aumentare la frequenza delle ricerche e, per via degli incontri ripetuti, la ritenzione [23]. Il beneficio non riguarda però la qualità del singolo atto di ricerca. Come accennato in §2.2, la componente search non risulta contribuire in modo significativo ai guadagni immediati e mostra persino un effetto lievemente negativo, seppur non significativo, su quelli differiti, in quanto lo sforzo di reperire l'informazione sottrarrebbe tempo e risorse cognitive alla codifica del termine [16]. A questo si aggiunge un limite strutturale indipendente dallo sforzo di consultazione dato dal fatto che la voce di dizionario è decontestualizzata per natura: elenca i significati possibili di un termine e lascia al lettore il compito di individuare quale si applichi al contesto specifico.

I traduttori automatici affrontano diversamente lo stesso problema. Questi traducono la porzione di testo selezionata anziché elencare significati, il che li rende efficaci su input estesi, ma su parole isolate o selezioni brevi restituiscono un'unica traduzione senza esporre le alternative, e i traduttori tradizionali commettono errori proprio sui termini polisemici, il cui significato dipende dal contesto [24]. La precisione migliora se il lettore amplia la selezione a più parole [25], spostando però l'onere della disambiguazione nuovamente su di lui.

Dizionari e traduttori integrati rimuovono dunque il primo vincolo del glossing tradizionale, la copertura limitata a un insieme fisso di parole, ma non il secondo: entrambi restituiscono un significato generico o una traduzione, non una spiegazione calibrata sul senso che il termine assume in quel testo. È su questo residuo, la spiegazione sensibile al contesto, che si collocano gli strumenti basati su modelli linguistici discussi nella sezione seguente.

### 2.4.3 Strumenti per la lettura in L2 basati su NLP e LLM

Gli strumenti basati su modelli linguistici di grandi dimensioni (LLM) intervengono proprio sul residuo lasciato aperto da dizionari e traduttori: invece di restituire una voce decontestualizzata o una traduzione, generano una spiegazione sensibile al contesto specifico, quando questo è fornito. Rispetto ai traduttori automatici tradizionali, questi modelli colgono meglio il significato dei termini polisemici e mantengono la coerenza del senso attraverso il testo, laddove la traduzione automatica convenzionale sbaglia proprio dove il significato dipende dal contesto [24]. Oltre alla traduzione, gli LLM possono spiegare anziché tradurre, sostenere un dialogo sul testo e adattare la formulazione al registro o al livello del lettore, tutte capacità che i sistemi precedenti non possedevano.

Su questa possibilità si innesta una linea di ricerca, in ambito NLP e HCI, dedicata a strumenti di supporto alla lettura per lettori non nativi. Già prima degli LLM, Azab et al. [26], descrivono uno strumento NLP in forma di e-book che offre accesso contestuale a più tipi di ostacolo, tra cui vocabolario sconosciuto, nomi non riconosciuti, frasi complesse e contesto perso tra una sessione e l'altra, pur operando su una libreria di testi pre-annotati.

Tra i sistemi qui esaminati, quello che più si avvicina a superare entrambi i limiti discussi finora è Reading.help [3], uno strumento LLM per lettori EFLarticolato in due moduli: un modulo di identificazione, basato su modelli NLP, che rileva parole e frasi potenzialmente difficili, e un modulo di spiegazione, basato su LLM, che chiarisce vocabolario, grammatica e contesto del passaggio. Il supporto che offre è insieme proattivo, poiché il sistema segnala il testo difficile, e su richiesta, potendo il lettore evidenziare qualunque porzione e chiederne la spiegazione.

Rispetto a una glossa pre-redatta o a una voce di dizionario, però, una spiegazione generata da LLM introduce un rischio nuovo: può essere errata. Nella valutazione di Reading.help le risposte risultano valide nell'88% dei casi per il vocabolario e nel 93% per la comprensione, con gli errori concentrati in allucinazioni e spiegazioni scorrette e una debolezza residua proprio sui termini a basso livello CEFR ma contestualmente insidiosi [3]. Gli strumenti basati su LLM rimuovono quindi i due vincoli dei sistemi precedenti, la copertura limitata a un insieme fisso di parole e la mancanza di calibrazione contestuale, ma ne introducono uno proprio, ovvero l'affidabilità della spiegazione generata.

# Capitolo 3

# Studio Formativo

Questo capitolo presenta lo studio formativo condotto per indagare i bisogni, le pratiche e gli ostacoli dei lettori italiani di inglese e raccogliere, parallelamente, il punto di vista di alcune esperte linguistiche. I risultati forniscono la base empirica per la successiva definizione dei requisiti dello strumento. Le sezioni seguenti descrivono le domande di ricerca, la metodologia adottata, il profilo del campione, i risultati emersi e i principali limiti dello studio.

## 3.1 Domande di ricerca

Lo studio formativo ha orientato la raccolta e l'analisi dei dati attorno a quattro domande di ricerca:

1. Cosa mette in difficoltà le persone quando leggono in inglese? Questo varia a seconda del livello di competenza linguistica?
2. La lettura è usata come occasione di apprendimento della lingua? Se sì, come?
3. Di cosa hanno bisogno le persone mentre leggono in inglese? Questi bisogni variano tra chi sta imparando la lingua e chi legge per altri scopi?
4. Come può la lettura supportare l'acquisizione linguistica e come questo processo può essere arricchito da un supporto esterno?

La sezione §3.4 presenta i risultati organizzati per domanda di ricerca, con una distinzione esplicita tra dati quantitativi del sondaggio e dati qualitativi delle interviste. La §3.4.4, dedicata alla quarta domanda, ha natura diversa: non presenta solo dati empirici raccolti direttamente, ma mette in dialogo i risultati delle sezioni precedenti con la letteratura SLA e con quanto emerso dalle interviste con le esperte linguiste.

## 3.2 Metodo

Lo studio ha adottato un disegno mixed-methods, che integra quindi dati quantitativi e qualitativi per rispondere a domande di ricerca complementari, articolato in tre fasi di raccolta dati: un sondaggio preliminare, interviste semistrutturate con lettori di inglese e interviste esplorative con esperte linguiste.

### 3.2.1 Partecipanti

Il reclutamento è avvenuto tramite passaparola, canali social e distribuzione di materiale informativo presso scuole di lingua e università di Torino. Quest'ultima modalità ha permesso di raggiungere anche apprendenti attivi che difficilmente sarebbero stati intercettati tramite i soli canali digitali. Non sono stati applicati criteri di esclusione predefiniti sulla frequenza di lettura o sullo status di apprendente: la scelta di includere sia chi studia attivamente la lingua sia chi legge in inglese per motivi non didattici è stata deliberata, nell'ottica di non restringere il campo d'indagine sulla base di assunzioni preliminari. Il sondaggio ha raccolto 89 risposte; sono state condotte 10 interviste con lettori di inglese e 3 con esperte linguiste.

### 3.2.2 Strumenti di raccolta dei dati

**Sondaggio** È stato distribuito un sondaggio online tramite Google Forms tra Gennaio e Febbraio 2026 che ha raccolto 89 risposte. Lo strumento comprendeva domande chiuse a scelta singola e multipla e due domande aperte, su temi relativi al profilo linguistico, alle abitudini di lettura, alle difficoltà incontrate durante la lettura in lingua inglese, ai comportamenti di apprendimento e agli strumenti attualmente utilizzati durante la lettura.

**Interviste con lettori di inglese.** Sono state condotte 10 interviste individuali semi-strutturate, della durata di circa 40 minuti ciascuna, in lingua italiana. Le interviste si sono svolte in videochiamata tramite Microsoft Teams. Ogni intervista prendeva avvio da un recap delle risposte fornite dal partecipante nel sondaggio, che fungeva da ancoraggio per approfondire quanto già dichiarato. La traccia tematica era articolata in: background di apprendimento della lingua; abitudini e contesti di lettura; approccio alla lettura in funzione del testo e dello scopo; difficoltà incontrate e loro tipologia; strategie adottate di fronte alle difficoltà; strumenti attualmente utilizzati; aspettative rispetto a uno strumento ideale. La conduzione è stata flessibile: le domande erano prevalentemente aperte, con follow-up adattativi, e la sequenza variava in funzione del corso della conversazione. Tutte le interviste includevano un episodio di lettura osservata: al partecipante veniva chiesto di

leggere un testo autentico in inglese, selezionato in base al livello CEFR autodichiarato, e di verbalizzare dubbi e difficoltà nel momento in cui si manifestavano. Data la variabilità dei testi, le osservazioni ricavate da questa fase sono trattate come dati qualitativi contestuali, utili ad approfondire e specificare quanto emerso nella parte discorsiva dell'intervista, e non come base per confronti sistematici tra partecipanti. Lo scopo del progetto veniva illustrato al termine dell'intervista, per non condizionare le risposte. Le interviste sono state registrate, previo consenso, e integralmente trascritte.

**Interviste con esperte linguiste** Parallelamente alle interviste con i lettori di inglese, sono state condotte 3 interviste semi-strutturate con esperte di linguistica applicata e didattica dell'inglese. Per tutelarne l'anonimato, le tre esperte intervistate sono qui indicate come E1, E2, E3: E1 è esperta linguistica presso il CLA del Politecnico di Torino; E2 è insegnante di inglese presso il CLA dell'Università di Torino; E3 è linguista e docente di lingua inglese presso l'Università di Torino. La traccia tematica comprendeva: difficoltà comuni degli studenti nell'apprendimento dell'inglese; ruolo della lettura nell'acquisizione linguistica; pratiche didattiche legate alla lettura; uso di strumenti digitali in contesto di apprendimento; consigli per lo studio autonomo. Come nelle interviste con i lettori di inglese, lo scopo del progetto veniva illustrato solo al termine, per non orientare le risposte verso una soluzione specifica; la chiusura raccoglieva impressioni e suggerimenti spontanei sullo strumento. Il loro ruolo è stato esplorativo. Da un lato hanno permesso di verificare se le difficoltà emergenti dal sondaggio e dalle interviste risuonassero con quanto osservato in contesti di insegnamento; dall'altro hanno offerto una prospettiva teorico-pratica sul ruolo della lettura nell'apprendimento linguistico. I loro contributi non costituiscono una conferma successiva dei risultati già raccolti, ma informano in modo secondario l'interpretazione dei dati e alcune delle implicazioni per il design discusse nella §3.4.4.

### 3.2.3 Approccio di analisi

I dati quantitativi del sondaggio sono stati analizzati con statistiche descrittive (frequenze, percentuali, distribuzioni per livello CEFR). I dati qualitativi delle interviste con lettori di inglese sono stati analizzati tematicamente, identificando pattern ricorrenti per categoria di difficoltà e livello di competenza. I dati delle esperte non sono stati confrontati sistematicamente con le altre due fonti, ma hanno orientato l'interpretazione dei risultati.

## 3.3 Profilo del campione

### 3.3.1 Sondaggio (n=89)

I rispondenti sono prevalentemente giovani adulti (età media 26,4 anni; range 18–62), con una netta concentrazione nella fascia 18–24 anni (n = 55, 62%). La quasi totalità ha dichiarato l'italiano come lingua madre. Sul piano dello status linguistico, 60 rispondenti (67%) non stanno studiando l'inglese attivamente ma lo leggono per motivi vari; 27 (30%) sono studenti attivi; 2 non rientrano in nessuna delle due categorie. La distribuzione per livello CEFR auto-dichiarato è la seguente:

**Tabella 3.1:** Distribuzione del campione per livello CEFR

| **Livello CEFR** | **n** | **%** |
|---|---|---|
| Non specificato | 2 | 2,2% |
| A1-A2 | 5 | 5,6% |
| B1 | 10 | 11,2% |
| B2 | 36 | 40,4% |
| C1-C2 | 36 | 40,4% |
| **Totale** | 89 | 100% |

I gruppi A1–A2 (n = 5) e B1 (n = 10) sono troppo piccoli per supportare conclusioni affidabili; le analisi disaggregate per livello si concentrano pertanto sui gruppi B2 e C1–C2. La frequenza di lettura è elevata: 49 rispondenti (55%) leggono in inglese ogni giorno, 16 (18%) più volte a settimana. I dispositivi più usati sono smartphone (n = 78) e computer (n = 72), con sovrapposizione frequente, seguiti poi da tablet (32), carta (26) e e-readers (10).

### 3.3.2 Interviste con lettori

Le 10 interviste hanno coperto i livelli da B1 a C2, con prevalenza di profili avanzati (5 intervistati C1–C2) e intermedi (3 B2, 2 B1). Quattro intervistati non studiano attivamente la lingua; sei sono apprendenti attivi, autonomi o formali. L'età è compresa tra i 22 e i 35 anni. I contesti d'uso prevalenti includono lettura accademica, professionale e di piacere.

## 3.4 Risultati

### 3.4.1 Difficoltà nella lettura in inglese

**Domanda di ricerca:** cosa mette in difficoltà le persone quando leggono in inglese? Questo varia a seconda dei livelli di competenza linguistica?

#### Dati del sondaggio

Le difficoltà incontrate durante la lettura in inglese si distribuiscono in tre categorie principali. Le espressioni idiomatiche risultano la difficoltà più diffusa (n = 52, 58%), seguite dal vocabolario individuale (n = 48, 54%) e dai riferimenti culturali o contestuali (n = 24, 27%). La grammatica è segnalata da un numero esiguo di rispondenti (n = 8, 9%), prevalentemente concentrati nel gruppo B1.

La distribuzione per livello CEFR mostra pattern differenziati, pur con la cautela che il gruppo B1 (n = 10) è troppo piccolo per supportare conclusioni affidabili:

**Tabella 3.2:** Difficoltà nella lettura per livello CEFR

| **Difficoltà** | **B1** ($n = 10$) | **B2** ($n = 36$) | **C1–C2** ($n = 36$) | **Totale** ($N = 89$) |
|---|---|---|---|---|
| Espressioni idiomatiche | 6 (60%) | 24 (67%) | 20 (56%) | 52 (58%) |
| Vocabolario individuale | 6 (60%) | 21 (58%) | 18 (50%) | 48 (54%) |
| Riferimenti culturali | 2 (20%) | 12 (33%) | 10 (28%) | 24 (27%) |
| Grammatica/ struttura frase | 2 (20%) | 2 (6%) | 0 (0%) | 8 (9%) |
| Tono/intenzione dell'autore | 0 (0%) | 3 (8%) | 3 (8%) | 7 (8%) |

Nota: Le percentuali sono calcolate sul totale del rispettivo gruppo CEFR. La domanda prevedeva risposta multipla; i valori non sommano a 100.

#### Dati delle interviste

Le interviste confermano e approfondiscono il quadro quantitativo, aggiungendo distinzioni qualitative che i soli dati del sondaggio non permettono di cogliere.

**Difficoltà lessicali.** La carenza di vocabolario emerge trasversalmente a tutti i livelli, ma con caratteristiche diverse a seconda del percorso di apprendimento e del livello di competenza. Questo è coerente con l'opinione di E1, secondo cui la

lacuna più diffusa tra gli studenti universitari italiani riguarda proprio il lessico, e in particolare, aggiunge, le collocazioni, ovvero sequenze di parole che ricorrono insieme e che vanno apprese come unità piuttosto che come termini isolati. Tra gli intervistati con un livello avanzato, le difficoltà lessicali riguardano principalmente termini formali e poco frequenti tipici della saggistica e del giornalismo, tecnicismi settoriali, espressioni metaforiche e termini arcaici, in generale un tipo di vocabolario che raramente si incontra al di fuori di certi registri scritti. Tra gli intervistati di livello intermedio e inferiore, in particolare chi ha appreso l'inglese prevalentemente in contesti scolastici, il vocabolario dei propri ambiti di specializzazione risulta spesso familiare, mentre le difficoltà maggiori riguardano parole di uso più quotidiano, slang (sia britannico che americano) ed espressioni che l'istruzione formale tende a non coprire. In entrambi i gruppi, inoltre, conoscere una parola non garantisce di saperne riconoscere il significato corretto nel contesto specifico in cui compare. Un aspetto trasversale ai livelli, segnalato sia nelle interviste che dalla letteratura, è la differenza tra comprensione ricettiva e uso attivo: molti partecipanti riferiscono di intuire il significato di un termine dal contesto senza però essere in grado di riutilizzarlo, il che suggerisce che la comprensione superficiale non si traduca automaticamente in acquisizione.

**Grammatica.** Le difficoltà grammaticali sono segnalate da un numero esiguo di partecipanti e riguardano quasi esclusivamente chi sta ancora acquisendo le strutture fondamentali della lingua. Nessun partecipante C1–C2 ha segnalato difficoltà grammaticali nella lettura; un'intervistata C1 ha precisato di incontrare difficoltà grammaticali solo nelle attività produttive, come il parlato. Un partecipante B2 ha espresso difficoltà non nella comprensione delle forme verbali in sé, ma nel capire perché in un determinato punto del testo venisse utilizzata una forma piuttosto che un'altra, quindi una difficoltà metalinguistica più che ricettiva. Questo dato è coerente con quanto osservato da E1, secondo cui alla fine della scuola dell'obbligo una base grammaticale consolidata è generalmente presente, mentre la carenza più diffusa riguarda il lessico.

**Difficoltà concettuali.** Una categoria di difficoltà distinta da quella strettamente lessicale riguarda la mancanza di conoscenze pregresse necessarie a comprendere il testo. In alcuni casi la difficoltà non è linguistica: anche traducendo correttamente un termine, la comprensione rimane parziale perché manca la conoscenza del contesto più ampio, che si tratti di un riferimento storico, politico o culturale non esplicitato nel testo, o semplicemente di un argomento con cui il lettore non ha familiarità. In questi casi non è sufficiente chiarire il significato di una parola ma è necessario approfondire l'argomento stesso. Un'intervistata ha esplicitato questo punto osservando che certe parole non le sarebbero chiare nemmeno se tradotte in italiano, perché il problema è concettuale prima che linguistico. Questa difficoltà

non è correlata al livello linguistico infatti la segnalano partecipanti di livelli diversi, ed è coerente con la distribuzione dei riferimenti culturali segnalati come tra le prime tre difficoltà in tutte le fasce CEFR.

**Monitoraggio della comprensione globale.** Trasversalmente ai livelli emerge una categoria di difficoltà distinta da quelle lessicali e culturali già documentate: l'incertezza sulla correttezza della propria comprensione globale di un passaggio. Diversi intervistati descrivono uno sforzo attivo di monitoraggio metacognitivo durante la lettura: verificare la coerenza tra ciò che si è capito e ciò che si sta leggendo, porsi domande su passaggi poco chiari, rileggere sezioni finché non si ritiene di averle comprese correttamente. Risulta interessante l'osservazione di una partecipante B1 che ricorre alla traduzione non come strumento di comprensione immediata, ma come strumento di verifica: traduce per controllare di aver capito, non per capire. Questo uso della traduzione segnala una difficoltà che non riguarda un elemento linguistico specifico, ma la capacità di monitorare e valutare la propria comprensione del testo nel suo insieme.

**Coerenza testuale.** Due dei tre intervistati B2 hanno segnalato difficoltà nel seguire il filo del discorso tra paragrafi, nel collegare idee quando il testo procede per salti tematici o quando non è chiaro a cosa si riferisca un termine o una frase. Si tratta di un dato numericamente marginale (3 persone contando anche il sondaggio), ma qualitativamente rilevante per la comprensione delle difficoltà specifiche al livello intermedio.

**Forma del testo.** In modo trasversale ai livelli, diversi intervistati hanno menzionato come ostacolo alla lettura non solo il contenuto linguistico ma la forma del testo che spesso si presenta con una struttura poco chiara, densità di informazione elevata e stile di scrittura poco accessibile. Questo aspetto non è catturato dalle categorie di difficoltà del sondaggio e emerge solo dal dato qualitativo.

### Sintesi

Le difficoltà nella lettura in inglese si articolano in tre categorie principali, coerenti tra dati quantitativi e qualitativi: lessicale, idiomatica e concettuale. Queste sono presenti a tutti i livelli di competenza, con espressioni idiomatiche e vocabolario individuale che risultano le difficoltà più diffuse nell'intero campione.

Le differenze per livello riguardano il tipo di vocabolario problematico più che la presenza o assenza della difficoltà lessicale in sé: i parlanti avanzati incontrano principalmente termini formali, rari o settoriali tipici di certi registri scritti; i parlanti intermedi e inferiori hanno maggiori difficoltà con vocabolario quotidiano,

slang e termini ambigui. Trasversale a tutti i livelli è la distanza tra comprensione ricettiva e uso attivo dei termini incontrati.

Le difficoltà concettuali, che includono riferimenti culturali non esplicitati ma anche più in generale la mancanza di familiarità con l'argomento, non sono correlate al livello linguistico e richiedono un tipo di supporto diverso dalla sola spiegazione lessicale.

La grammatica emerge come difficoltà solo per chi è ancora in fase di acquisizione delle strutture fondamentali, e scompare tra i parlanti avanzati.

Una quarta categoria, emersa esclusivamente dal dato qualitativo, riguarda il monitoraggio della comprensione globale: l'incertezza sulla correttezza della propria interpretazione di un passaggio, indipendentemente dalla presenza di difficoltà lessicali o culturali specifiche.

Un altro aspetto emerso esclusivamente dal dato qualitativo riguarda la forma del testo: struttura poco chiara, densità informativa elevata e stile di scrittura ostico vengono citati come ostacoli alla lettura indipendentemente dal livello linguistico.

### 3.4.2 La lettura come occasione di apprendimento

**Domanda di ricerca:** La lettura è usata come occasione di studio della lingua? Se sì, come?

#### Dati del sondaggio

Circa la metà del campione (n = 45, 51%) dichiara di adottare comportamenti orientati all'apprendimento linguistico durante la lettura, pur non avendolo come obiettivo primario. L'apprendimento tramite lettura è quindi un fenomeno diffuso ma prevalentemente opportunistico, poiché accade mentre si legge per altri scopi, non come attività deliberatamente didattica. Il comportamento più comune a tutti i livelli è reattivo, ovvero ci si ferma su una parola sconosciuta e la si cerca. È una strategia presente dall'A1–A2 al C1–C2 e da sola non distingue chi apprende in modo sistematico da chi adotta comportamenti di apprendimento più occasionali. Con l'aumentare della competenza le strategie si diversificano: ai livelli bassi prevale la decodifica tramite traduttore o dizionario; ai livelli intermedi compaiono la presa di note sistematica, la ripetizione ad alta voce e il tentativo di uso attivo delle parole apprese; ai livelli avanzati emergono comportamenti più elaborati, come l'uso di dizionari monolingui, l'analisi di sinonimi e contesti d'uso e l'integrazione dell'inglese nella routine quotidiana. Quattordici rispondenti (16%), equamente distribuiti tra studenti formali e autodidatti, hanno dichiarato di scegliere intenzionalmente testi impegnativi per migliorare dal punto di vista linguistico. La distinzione più rilevante rispetto ai comportamenti di apprendimento non è tra chi studia

formalmente e chi no, le proporzioni delle strategie adottate sono infatti quasi identiche nei due gruppi, ma tra livelli di competenza.

### Dati delle interviste

Nessun intervistato dichiara di leggere con l'obiettivo primario di imparare l'inglese, con la parziale eccezione dei partecipanti B1 impegnati nella preparazione all'esame di lingua "IELTS", per i quali comunque la lettura è funzionale alla certificazione, non all'apprendimento della lingua in sé. L'apprendimento tramite lettura emerge come effetto collaterale o come processo riconosciuto a posteriori.

Nei partecipanti C1–C2 il meccanismo è prevalentemente implicito e non mediato da strumenti. Le parole vengono incontrate, cercate se necessario e trattenute per esposizione ripetuta, non per studio deliberato; l'annotazione sistematica è assente o sporadica. Un partecipante descrive il proprio percorso di apprendimento, avvenuto attraverso film e lettura, senza studio formale, come analogo a quello di un madrelingua: le regole vengono applicate correttamente pur senza essere ricordate in modo esplicito. Un altro identifica i social media come canale principale di acquisizione dello slang e dei registri informali, e teorizza esplicitamente l'esposizione ripetuta come il modo in cui si "impara davvero una lingua". Lo stesso partecipante riconosce però che la lettura di testi accademici e professionali in università ha svolto una funzione specifica nell'acquisizione di registri formali che la conversazione informale non avrebbe potuto fornire, una distinzione che attribuisce consapevolmente alla lettura come canale di apprendimento.

Nei partecipanti B2 il quadro è più frammentato. Un partecipante utilizza piattaforme come LinkedIn e Duolingo come canali di esposizione quotidiana alla lingua, riconoscendo di aver acquisito vocabolario in contesto, ma senza annotazione o ripasso sistematico. L'apprendimento esiste ma è destrutturato e non monitorato. Un altro partecipante tende ad abbandonare il testo quando il vocabolario è troppo denso, interrompendo l'esposizione prima che qualsiasi processo di apprendimento possa verificarsi.

Nei partecipanti B1 il meccanismo è opposto: esplicito, intenzionale e mediato da strumenti. Le parole vengono annotate sistematicamente su supporti fisici e digitali. Un partecipante osserva che il vocabolario non viene memorizzato in modo isolato ma si organizza spontaneamente per settori semantici, una struttura cognitiva che emerge dall'uso, non da una tecnica di studio deliberata. Questo comportamento sistematico si manifesta però principalmente nel contesto del corso formale: nella lettura autonoma i comportamenti diventano più reattivi e meno strutturati.

Inoltre, nei partecipanti di livello intermedio e inferiore risulta esserci una tensione tra il desiderio di usare la lettura come occasione di apprendimento e la difficoltà dei testi autentici. Infatti, quando il carico lessicale percepito diventa

troppo elevato, alcuni partecipanti riferiscono di interrompere la lettura o di orientarsi verso testi più accessibili.

#### Sintesi

L'apprendimento linguistico emerge durante la lettura come esito secondario e opportunistico, più che come obiettivo primario dell'attività. In particolare, avviene quando si incontra una difficoltà o per effetto dell'esposizione ripetuta, non come scopo della lettura stessa. La distinzione che appare più rilevante rispetto alle modalità di apprendimento non è tra chi studia formalmente e chi è autodidatta, in quanto le strategie adottate sono simili nei due gruppi, ma tra livelli di competenza.

Nei livelli avanzati il processo è implicito, basato sull'esposizione ripetuta e non mediato da strumenti sistematici; la lettura di testi formali viene riconosciuta retrospettivamente come canale specifico per l'acquisizione di registri che altri contesti non forniscono. Nei livelli intermedi il processo è destrutturato e discontinuo: l'apprendimento avviene ma non viene monitorato né consolidato. Nei livelli bassi il processo è intenzionale e mediato da strumenti, ma si manifesta principalmente nel contesto formale; nella lettura autonoma è ostacolato dalla difficoltà di trovare testi al livello giusto senza supporto esterno.

### 3.4.3 Bisogni durante la lettura in inglese

**Domanda di ricerca:** Di cosa hanno bisogno le persone mentre leggono in inglese? Questi bisogni sono diversi per chi sta imparando la lingua?

#### Dati del sondaggio

Alla domanda su cosa vorrebbero poter chiedere durante la lettura, i rispondenti indicano come priorità: definizione o traduzione dei termini (n = 50, 56%), esempi di utilizzo in contesti diversi (n = 34, 38%), informazioni culturali o contestuali (n = 26, 29%), spiegazione in inglese più semplice (n = 24, 27%), confronto con parole o strutture simili (n = 20, 22%). Più in basso si collocano la spiegazione del perché l'autore ha scelto certe parole (n = 9, 10%), la spiegazione grammaticale (n = 9, 10%) e gli aiuti visivi (n = 14, 16%). Solo 6 rispondenti (7%) dichiarano di non aver bisogno di nulla.

Una parte del sondaggio indagava anche i comportamenti di evidenziazione del testo. Trentasette rispondenti dichiarano di non evidenziare affatto, mentre, tra chi lo fa, le motivazioni principali sono ricordare informazioni importanti (n = 28), marcare parole da cercare dopo (n = 19) e copiare testo da tradurre (n = 17); solo 8 evidenziano per identificare parti confuse e 5 per analizzare grammatica o struttura.

**Tabella 3.3:** Bisogni durante la lettura in inglese

| **Bisogno** | **n** | **%** |
|---|---|---|
| Definizione/traduzione | 50 | 56 |
| Esempi di utilizzo in contesti diversi | 34 | 38 |
| Informazioni culturali/contestuali | 26 | 29 |
| Spiegazione in inglese più semplice | 24 | 27 |
| Confronto con parole/strutture simili | 20 | 22 |
| Aiuti visivi (immagini, diagrammi) | 14 | 16 |
| Spiegazione grammaticale | 9 | 10 |
| Spiegazione delle scelte dell'autore | 9 | 10 |
| Nessun bisogno | 6 | 7 |

Nota: la domanda prevedeva risposta multipla; i valori percentuali sono calcolati sul totale dei rispondenti ($n = 89$) e non sommano a 100.

### Dati delle interviste

Il bisogno principale emerso dalle interviste è risolvere i dubbi durante la lettura in modo immediato e contestualmente accurato, senza interrompere il flusso. La frustrazione più ricorrente riguarda l'inaccuratezza contestuale degli strumenti attuali: le risposte vengono descritte come troppo letterali, incapaci di cogliere il senso idiomatico o di adattarsi al contesto specifico del testo. A questo si aggiunge il fastidio di dover ricorrere a strumenti esterni, con il rischio di perdere il filo della lettura o distrarsi. Un'intervistata cita l'integrazione del dizionario nel Kindle come esempio di interazione apprezzata proprio perché non richiede di uscire dal testo, pur riconoscendone i limiti rispetto a espressioni idiomatiche e riferimenti culturali.

Alcuni intervistati riferiscono di marcare i termini problematici per cercarne il significato a lettura terminata, piuttosto che interrompere immediatamente. Quando i dubbi vengono risolti, la preferenza è di ritrovare la spiegazione direttamente nel testo, perché rileggendo lo stesso passaggio in un secondo momento il significato cercato viene spesso dimenticato.

Il bisogno di esempi di utilizzo in contesti diversi è espresso con chiarezza da partecipanti di livello B1 e B2: vedere come un termine viene usato in casi diversi non solo chiarisce il passaggio corrente, ma permette di capire in quali contesti può essere riutilizzato. Diversi partecipanti dicono inoltre di memorizzare i termini associandoli al contesto di ritrovamento piuttosto che alla traduzione isolata.

Una parte dei partecipanti esprime una preferenza esplicita per ricevere spiegazioni nella lingua target, per non rompere l'immersione linguistica e per favorire una comprensione del termine nel suo contesto naturale.

Le spiegazioni grammaticali contestualizzate emergono come bisogno circoscritto:

presente tra alcuni rispondenti al sondaggio di livello B2 e tra chi studia attivamente la lingua, assente tra i livelli avanzati. Un intervistato B2 esprime il bisogno di capire non la forma grammaticale in sé, ma il motivo per cui viene usata in un determinato punto del testo.

Due intervistate C1-C2 esprimono un bisogno che va oltre la risoluzione di dubbi linguistici: la possibilità di dialogare con il testo a livello concettuale, approfondire argomenti e comprendere le scelte stilistiche e comunicative dell'autore.

Un bisogno di natura diversa, e trasversale ai livelli a differenza del precedente, riguarda la verifica della comprensione globale di un passaggio. Alcuni partecipanti descrivono strategie di compensazione spontanee quali rileggere più volte o ricorrere alla traduzione come controllo, che segnalano non una difficoltà con un termine specifico, ma l'incertezza sulla correttezza della propria interpretazione complessiva. Questo bisogno emerge prevalentemente in chi legge per estrarre informazioni o completare un compito, e appare indipendente dal livello linguistico.

Tre partecipanti su dieci esprimono una preferenza per l'interazione vocale come modalità di risoluzione dei dubbi durante la lettura. Il caso più diretto è quello di una partecipante C1 che descrive lo strumento ideale come un assistente a cui parlare senza dover scrivere, motivando esplicitamente la preferenza con la riduzione dell'attrito: poter fare una domanda a voce nel momento in cui emerge il dubbio, senza interrompere la lettura per digitare. Un partecipante B2 esprime una preferenza analoga per un tutor vocale interattivo, in un contesto però più ampio dell'apprendimento tramite lettura. Una partecipante B1 immagina uno strumento sempre presente e portabile a cui rivolgersi spontaneamente, suggerendo implicitamente una modalità vocale. Negli altri sette partecipanti questa preferenza non emerge: chi descrive l'attrito con gli strumenti attuali immagina soluzioni testuali integrate nel testo, non vocali.

Un interesse a ricevere aiuti visivi è espresso da 14 rispondenti al sondaggio(16%). Nelle interviste, un partecipante descrive come nella lettura di narrativa o testi scientifici un'illustrazione possa facilitare la comprensione più di una spiegazione testuale.

Più partecipanti menzionano l'uso spontaneo di contenuti video con sottotitoli come strategia di apprendimento. Il comportamento suggerisce una preferenza per l'associazione tra forma scritta e forma orale, anche se non viene espresso come bisogno esplicito nel contesto della lettura di testi.

### Sintesi

Il bisogno trasversale a tutti i livelli è ricevere supporto contestualmente accurato senza interrompere il flusso di lettura. La traduzione immediata è il bisogno più diffuso e il più veloce da soddisfare, ma è percepita come insufficiente da chi vuole andare oltre la comprensione momentanea. Attorno a questo nucleo si

distribuiscono bisogni più specifici che variano per livello e per orientamento verso l'apprendimento, quali esempi d'uso, spiegazioni in L2, informazioni culturali e contestuali e sinonimi. Le spiegazioni grammaticali contestualizzate e il dialogo con il testo a livello concettuale emergono come bisogni di nicchia, rispettivamente ai livelli intermedi e avanzati. Chi studia attivamente la lingua tende a esprimere richieste più variate rispetto a chi legge per altri scopi, indipendentemente dal livello di competenza. Un bisogno trasversale ai livelli, emerso esclusivamente dal dato qualitativo, riguarda la verifica della comprensione globale di un passaggio, indipendentemente dalla presenza di difficoltà lessicali specifiche. Una preferenza per l'interazione vocale come riduzione dell'attrito nella risoluzione dei dubbi emerge come bisogno secondario e minoritario, presente in tre interviste su dieci e collegato esplicitamente alla lettura solo in un caso. Un bisogno comune riguarda infine la gestione delle incomprensioni nel tempo: molti partecipanti preferiscono ritrovare le spiegazioni direttamente nel testo piuttosto che in uno strumento separato.

### 3.4.4 Come la lettura può supportare l'acquisizione linguistica

Questa sezione ha natura diversa dalle precedenti in quanto non risponde a una domanda empirica attraverso dati raccolti, ma mette in dialogo i risultati dello studio con la letteratura SLA e con quanto emerso dalle interviste con le esperte.

#### Cosa emerge dai dati

I risultati delle sezioni precedenti delineano un quadro in cui l'acquisizione linguistica avviene prevalentemente in forma incidentale, così come definita in §2.1. La maggior parte dei partecipanti non legge con l'obiettivo di imparare la lingua, ma circa la metà adotta comportamenti orientati all'apprendimento quando incontra una difficoltà: cercando il significato di un termine, annotandolo, cercando di memorizzarlo nel contesto di ritrovamento. L'apprendimento è quindi secondario, dipende dall'incontro con un ostacolo e dalla disponibilità cognitiva e motivazionale di affrontarlo in quel momento.

I comportamenti osservati tra gli intervistati riflettono la distinzione tra apprendimento incidentale e intenzionale discussa in §2.1. I partecipanti avanzati si affidano prevalentemente all'esposizione ripetuta, propria dell'apprendimento incidentale, mentre i partecipanti intermedi e bassi mostrano tentativi più deliberati come annotare, cercare, contestualizzare, riconducibili a un approccio più intenzionale, che però raramente vengono consolidati attraverso la riattivazione nel tempo. Questo è coerente con quanto discusso in §2.2: la ripetizione da sola non garantisce un apprendimento solido se non coinvolge recupero attivo e incontro variato, condizioni che l'annotazione isolata, senza ripasso, non soddisfa.

Due tensioni emergono dai dati. La prima riguarda la traduzione, lo strumento più usato e il più rapido per risolvere dubbi, ma percepita come insufficiente da chi vuole andare oltre la comprensione momentanea; i partecipanti stessi riconoscono di dimenticare facilmente i termini cercati in questo modo. La seconda riguarda chi si avvicina a testi autentici con un livello ancora insufficiente. Il desiderio di apprendere tramite la lettura è presente, ma la difficoltà dei testi funziona come deterrente, riducendo l'esposizione alla lingua prima che qualsiasi apprendimento possa avvenire, un effetto coerente con la soglia di tolleranza discussa in §2.3

**Interpretazione teorica e contributo delle esperte**

La tensione sulla traduzione trova una spiegazione diretta nell'Involvement Load Hypothesis discussa in §2.2: usata come unica modalità di risoluzione del dubbio, la traduzione attiva il need ma non la valutazione contestuale, la componente che la ricerca sull'ILH conferma come più determinante per la ritenzione. E3 offre una lettura convergente da una prospettiva didattica, distinguendo tra scopi comunicativi e scopi di acquisizione: la traduzione risolve rapidamente un dubbio, ma richiede un'elaborazione cognitiva insufficiente per consolidare il termine in memoria a lungo termine. Questo non è un argomento contro la traduzione in quanto tale, né contro la lingua in cui è formulata. La ricerca sui gloss in L1 e L2 discussa in §2.4.1 spesso mostra anzi un vantaggio della L1 per la ritenzione, moderato dal livello del lettore. Il punto sollevato dai dati e da E3 riguarda piuttosto il livello di elaborazione che l'attività richiede, indipendentemente dalla lingua. Per la comprensione immediata la traduzione resta lo strumento più efficace, mentre per l'acquisizione a lungo termine serve un'elaborazione più attiva.

La seconda tensione, quella sui testi troppo difficili, trova nella convergenza tra le esperte e la letteratura sulla lettura estensiva (§2.3) una spiegazione complementare. E1 e E2 convergono sull'importanza della motivazione e del piacere della lettura come condizioni necessarie per l'apprendimento, osservando che trasformare la lettura in una dissezione sistematica del testo rischia di ridurre la motivazione intrinseca, vanificando i benefici dell'esposizione. A questo si aggiunge, sul piano linguistico, quanto discusso in §2.3, ovvero che un carico lessicale sopra una certa soglia porta il lettore ad abbandonare il testo o a smettere di elaborarlo attivamente, indipendentemente da come l'attività di lettura sia proposta.

Il principio dell'elaborazione attiva emerso a proposito della traduzione si estende, secondo le tre esperte, a qualunque interazione con il testo. E3 segnala che il dialogo mediato da uno strumento attiva una forma di metacognizione: formulare una domanda su un termine richiede di elaborarlo attivamente, aumentando il coinvolgimento cognitivo e la probabilità di ritenzione. E2 descrive una dinamica funzionalmente diversa ma convergente nell'obiettivo: nelle sue attività di lettura guidata è l'insegnante a porre domande agli studenti sul testo, orientandone

l'attenzione e chiedendo di localizzare le risposte nel testo stesso. E1 sottolinea più in generale l'importanza di mantenere gli studenti attivamente coinvolti, osservando che qualsiasi cosa li tolga dalla partecipazione attiva al processo non li aiuta ad apprendere. Che sia l'utente a formulare una domanda, lo strumento a proporne una, o il processo di lettura stesso a richiedere un'elaborazione attiva, l'interazione consapevole con il testo produce un livello di elaborazione più profondo rispetto alla sola comprensione passiva.

Su un piano diverso, E3 segnala inoltre che l'input audio è più impattante di quello visivo per la ritenzione lessicale: sentire un termine oltre a vederlo scritto aumenta la probabilità che venga consolidato in memoria, suggerendo che una componente audio non sarebbe un arricchimento accessorio ma un supporto con una giustificazione pedagogica propria.

Questi elementi convergono verso alcuni principi che orientano il design dello strumento: il supporto alla comprensione dovrebbe integrarsi nel flusso della lettura senza imporsi ad essa e dovrebbe offrire una risposta minima immediata, non necessariamente solo testuale, con possibilità di approfondimento, lasciando all'utente la scelta di quanto investire in ciascun incontro con una difficoltà. Dovrebbe inoltre ridurre la barriera lessicale che allontana i lettori da testi altrimenti troppo difficili, preservando l'esposizione necessaria perché l'apprendimento incidentale possa avvenire. La traduzione rimane uno strumento legittimo e necessario per la comprensione immediata, ma non sufficiente da sola per l'acquisizione lessicale, non per la lingua in cui è formulata, bensì per il livello di elaborazione cognitiva che richiede rispetto ad attività che coinvolgono valutazione contestuale attiva. I requisiti che derivano da questi principi sono presentati in (§4).

### 3.4.5 Sintesi trasversale

Attraverso le quattro domande di ricerca emergono alcuni pattern ricorrenti che attraversano i dati in modo trasversale e che definiscono il contesto entro cui lo strumento deve operare.

La difficoltà lessicale è presente a tutti i livelli di competenza, ma non è uniforme. I partecipanti avanzati incontrano principalmente vocabolario formale e poco frequente tipico di certi registri scritti; i partecipanti intermedi hanno difficoltà sia con vocabolario quotidiano e slang, che il percorso scolastico tende a non coprire, sia con termini il cui significato nella traduzione isolata non è sufficiente a chiarire come vengono usati nel contesto specifico del testo. Questo suggerisce che un supporto efficace debba essere adattabile al contesto piuttosto che standardizzato.

L'apprendimento linguistico è un obiettivo secondario e non pianificato, ma presente in circa la metà del campione. Lo strumento non deve imporlo, poiché farlo rischierebbe di compromettere la motivazione alla lettura, ma non può nemmeno

ignorarlo, dato che una parte significativa degli utenti cercherà attivamente di sfruttare l'incontro con una difficoltà come occasione di apprendimento. Lo strumento dovrebbe quindi assecondare questo comportamento opportunistico senza imporlo, abbassando il costo dell'approfondimento nel momento in cui emerge il bisogno, senza trasformare la lettura in un'attività didattica strutturata.

Il bisogno principale è la risoluzione contestualmente accurata dei dubbi senza interrompere il flusso di lettura. Gli strumenti attuali soddisfano questo bisogno solo parzialmente: o sono integrati nel testo ma poco accurati nel contesto, o sono accurati ma richiedono di uscire dal testo. Questa lacuna è trasversale ai livelli e ai profili.

Emerge una tensione strutturale tra comprensione immediata e acquisizione a lungo termine. Quello che è più efficace per risolvere rapidamente un dubbio, ovvero la traduzione, è meno efficace per consolidare un termine in memoria. Uno strumento che privilegia la velocità soddisfa il bisogno primario ma non quello secondario; uno che privilegia l'approfondimento rischia di diventare intrusivo. Questa tensione non si risolve eliminando uno dei due poli, ma costruendo una gerarchia di supporto che li renda entrambi accessibili su richiesta, offrendo spiegazioni direttamente invece di richiedere all'utente lo sforzo di cercarle, dato che è la qualità dell'elaborazione a favorire la ritenzione e non la ricerca in sé.

La distinzione più rilevante tra profili non è tra chi studia formalmente la lingua e chi non la studia, ma tra livello di competenza e orientamento verso l'apprendimento. Questi due assi producono profili di bisogno distinti che emergono chiaramente dai dati e che non coincidono con la distinzione tra apprendente e non apprendente.

Infine, emerge una tensione specifica che riguarda chi si avvicina a testi autentici con un livello ancora insufficiente. La motivazione a imparare tramite lettura è presente, ma la difficoltà dei testi funziona come deterrente, riducendo l'esposizione alla lingua prima che qualsiasi apprendimento possa avvenire. Questa tensione rimane irrisolta nei dati e rappresenta una delle sfide progettuali più rilevanti per i profili di livello più basso.

Questi pattern convergono verso tre profili di utente distinti, che si differenziano per livello di competenza, orientamento verso l'apprendimento e tipo di supporto ricercato durante la lettura. Le personas sviluppate nel capitolo seguente istanziano questi profili a partire dai dati raccolti, e costituiscono il punto di partenza per la definizione dei requisiti dello strumento.

## 3.5 Limiti dello studio formativo

I risultati riportati vanno interpretati con cautela, alla luce di alcuni limiti che caratterizzano lo studio formativo condotto.

Il livello di competenza linguistica dei partecipanti, sia nel sondaggio sia nelle interviste, è dichiarato dai partecipanti stessi, sulla base di certificazioni possedute o di autovalutazione, ma non verificato tramite un test di livello somministrato dallo studio. Le distribuzioni per fascia CEFR riportate in questo capitolo vanno quindi lette come una stima non controllata direttamente e non come una misura oggettiva, ma indicativa, della competenza dei partecipanti.

Il reclutamento tramite canali universitari e passaparola ha prodotto un campione concentrato in una fascia d'età ristretta (età media 26,4 anni, 62% tra i 18 e i 24 anni) e con un livello di istruzione presumibilmente elevato, limitando la generalizzabilità dei risultati a fasce d'età più ampie o a contesti extra-universitari. I gruppi di livello A1-A2 e B1 sono inoltre troppo piccoli (rispettivamente $n = 5$ e $n = 10$) per supportare conclusioni affidabili, una limitazione rilevante dato che la tensione tra motivazione e difficoltà dei testi, discussa in precedenza, riguarda proprio i lettori di livello più basso, per i quali i dati raccolti sono i più esigui.

L'episodio di lettura osservata condotto durante le interviste ha utilizzato testi diversi per ciascun partecipante, selezionati in base al livello CEFR autodichiarato ma non calibrati individualmente sul suo vocabolario noto. Per la maggior parte dei partecipanti questo non ha prodotto osservazioni particolarmente rilevanti, sia perché il testo assegnato non presentava difficoltà sufficienti a generare dubbi osservabili, sia perché il contesto stesso dell'osservazione, con un brano assegnato, letto durante un'intervista in videochiamata e sapendo di essere osservati, si discosta dalle condizioni della lettura spontanea che lo studio intende indagare.

L'analisi tematica delle interviste è stata condotta da un'unica persona, senza il supporto di un secondo codificatore con formazione specifica in linguistica applicata, il che può aver influenzato l'interpretazione di alcune categorie. La stessa persona ha inoltre ricoperto contemporaneamente il ruolo di intervistatore e di annotatore, contro la prassi raccomandata di separare questi due ruoli per ridurre il carico cognitivo e il rischio di orientare inconsapevolmente le risposte.

Più in generale, il numero di partecipanti coinvolti, in particolare nelle interviste, resta troppo esiguo per garantire una generalizzabilità statistica dei risultati a tutta la popolazione di lettori italiani di inglese. Questo è in parte coerente con la natura esplorativa dello studio formativo, che punta a identificare pattern qualitativi e bisogni ricorrenti più che a stimarne la prevalenza esatta nella popolazione; tuttavia, poiché questi stessi risultati sono stati usati per orientare scelte di design vincolanti per l'intero prototipo, sono comunque presenti margini di incertezza da tenere in considerazione qualora tali scelte venissero riesaminate con un campione più ampio o diversificato.

# Capitolo 4

# Definizione dei Requisiti

Questo capitolo traduce i risultati dello studio formativo in una base concreta per il design dello strumento. A partire dalle personas e dai context scenarios, i bisogni degli utenti vengono sintetizzati e sottoposti a un processo di definizione delle priorità, da cui discendono infine i requisiti funzionali e non funzionali del sistema.

## 4.1 Personas e Context Scenarios

Questa sezione presenta le personas sviluppate a partire dai dati dello studio formativo e i context scenarios che ne derivano, secondo l'approccio del Goal-Directed Design descritto in §1.2

### 4.1.1 Definizione delle personas

Le informazioni raccolte durante lo studio formativo sono state usate per la creazione di tre personas, incarnazioni di comportamenti e necessità osservate che accomunano i vari partecipanti e i potenziali utenti dello strumento, utili a facilitare la fase di design.

Ogni persona non rappresenta un singolo individuo nella sua totalità, ma incarna comportamenti e bisogni ricorrenti tra i partecipanti. Vale la pena notare che alcuni partecipanti mostrano comportamenti riconducibili a più di una persona a seconda del contesto. Le personas rappresentano quindi tendenze dominanti, non categorie rigide ed esclusive.

**Laura - L'esperta tecnica** Laura è una studentessa universitaria di 25 anni che ha raggiunto un alto livello di competenza linguistica attraverso corsi formali e periodi di studio all'estero. Non studia più l'inglese e non sente il bisogno di approfondirne ulteriormente la conoscenza: per lei è un mezzo più che un fine, che le

permette di leggere paper scientifici e reperire informazioni difficilmente disponibili in italiano. Legge principalmente per studio o ricerca, sempre al computer, con l'obiettivo di estrarre informazioni precise e comprendere concetti complessi.

A metterla in difficoltà sono parole rare, di cui cerca prima la definizione in inglese poi la traduzione se necessario, e termini tecnici legati all'argomento, che approfondisce sul web o con un LLM. Risolvere questi dubbi le rallenta il lavoro, quindi spesso posticipa la ricerca a dopo la lettura. Raramente le incomprensioni sono abbastanza numerose da ostacolare la comprensione complessiva del testo. Evidenzia regolarmente parti di testo per ritrovarle e integrarle nel suo lavoro.

**Bianca - L'appassionata autodidatta** Bianca è una giovane lavoratrice di 26 anni che ha imparato l'inglese in parte a scuola, ma principalmente attraverso un approccio autonomo di immersione nella lingua, che le ha permesso di raggiungere un livello avanzato. Si sente abbastanza sicura nella comprensione di slang e parole d'uso quotidiano, ma riscontra difficoltà con il vocabolario formale o accademico, i termini specifici degli argomenti che legge e i riferimenti culturali o storici a lei sconosciuti. Legge in inglese per piacere, saggistica e narrativa, sia per mantenersi allenata nella lingua sia per cogliere le sfumature dei testi nella loro lingua originale. La incuriosisce scoprire l'origine di espressioni idiomatiche o il significato culturale dietro certi usi della lingua, e vorrebbe poter dialogare con i testi e interrogarne gli autori. Conta sulla lettura come occasione principale per interiorizzare gradualmente termini ed espressioni, non avendo occasioni regolari di usare l'inglese in contesti produttivi. Quando incontra qualcosa che non capisce, spesso va avanti ignorando l'incomprensione pur di non interrompere la lettura. È poi spesso costretta a tornare sui termini poco chiari dopo aver realizzato di non aver capito il significato generale del testo. Approfondisce quando ne ha voglia, cercando definizioni in lingua o informandosi sul contesto, ma farlo richiede di uscire dal testo e raramente è disposta a farlo.

**Mattia - L'apprendente attivo** Mattia ha 22 anni ed è uno studente universitario che studia l'inglese attivamente, sia da autodidatta sia attraverso corsi formali, in preparazione di una certificazione linguistica. È abbastanza a suo agio con i testi inerenti al proprio ambito di specializzazione, ma si sente molto in difficoltà con il lessico comune e fatica a comprendere testi su temi di vita quotidiana o ambiti al di fuori dei suoi interessi. Le espressioni idiomatiche e lo slang lo confondono perché non riesce a dedurne il significato dalla traduzione letterale. Non gli è sempre chiaro il ragionamento dietro l'uso di certe forme verbali o sintattiche. Legge articoli e blog su argomenti che lo interessano, oltre ai testi forniti a lezione. Quando incontra un termine nuovo lo appunta con sinonimi ed esempi d'uso, e ogni tanto ripassa le liste sperando di interiorizzare i termini nuovi. Ricorre principalmente a traduttori online come Google Translate o DeepL, ma sente che spesso non restituiscono il

significato più appropriato al contesto, a meno che non inserisca l'intera frase o paragrafo.

### 4.1.2 Definizione dei context scenarios

I context scenarios descrivono come ciascuna persona interagisce con un sistema ideale, capace di rispondere ai propri bisogni, in un contesto d'uso rappresentativo. Seguendo l'approccio del *magic thinking* proposto da Cooper et al.[7], gli scenari immaginano un sistema che risponde esattamente ai bisogni dell'utente nel momento in cui emergono, senza che l'utente debba adattarsi a vincoli tecnici o di implementazione. I needs derivati da questi scenari costituiranno la base per la definizione dei requisiti. Non tutte le funzionalità descritte negli scenari sono state incluse nel prototipo: i criteri con cui ad alcune è stata data la priorità e altre rinviate a sviluppi futuri sono esplicitati nella §4.2.

**Laura - Documentarsi per la tesi** Laura deve documentarsi su un argomento per la tesi. Si siede alla scrivania e apre un paper sul computer: sa che avrà bisogno di uno schermo grande e di tenere aperte più applicazioni contemporaneamente. Mentre legge incontra una parola inusuale. Sa che potrebbe cercarla, ma farlo significa aprire un'altra scheda e perdere il filo. L'interfaccia le propone un termine equivalente che già conosce e che è appropriato al contesto: Laura capisce e prosegue senza essersi fermata davvero.

Poco dopo incontra un termine tecnico che non conosce, un concetto nuovo, centrale per capire il passaggio. Rilegge la frase, ma il contesto da solo non basta. Di solito andrebbe avanti sperando che il senso generale emergesse dopo, ma questa volta il termine è troppo importante per ignorarlo. L'interfaccia le fornisce una spiegazione contestualizzata che chiarisce il ruolo del termine nel testo. Laura non è del tutto sicura di aver capito correttamente e lo comunica all'interfaccia, che conferma la sua interpretazione. Soddisfatta, sceglie di andare avanti.

Più avanti si trova davanti a un paragrafo denso che richiederebbe di rallentare e concentrarsi. Non vuole perdere lo slancio della lettura, quindi lo marca per tornarci dopo. Quando ci ritorna, l'interfaccia la guida nell'individuare i punti chiave e nel verificare la sua comprensione man mano.

A lettura conclusa, l'interfaccia ha raccolto tutte le informazioni che Laura ha marcato come rilevanti durante la lettura, pronte per essere recuperate e integrate nel lavoro che sta scrivendo.

**Bianca - Leggere per piacere** Bianca è sul divano la sera, sta leggendo un saggio in inglese sul suo tablet.

Ogni tanto incontra parole che non le sono chiare. Di solito valuterebbe se vale la pena fermarsi o se il senso generale rimane comunque comprensibile, e spesso

sceglie di andare avanti pur di non dover aprire un'altra app. Stavolta l'interfaccia è lì, e il costo dell'interruzione scompare: seleziona la parola e ottiene una spiegazione contestualizzata senza uscire dal testo. A volte le vengono suggeriti altri termini che potrebbero essere usati per esprimere lo stesso concetto.

A un certo punto il testo fa riferimento a un fatto storico di cui non è a conoscenza. Perde il filo del discorso e per un momento considera di andare avanti lo stesso. Poi decide di no: capire quel riferimento è necessario per seguire il ragionamento dell'autore. L'interfaccia le spiega di cosa si tratta, chi sono gli attori coinvolti e perché è rilevante in quel punto del testo.

Arrivata a un paragrafo particolarmente complesso rallenta. Terminata la lettura, ha una comprensione vaga dell'argomentazione ma sente che le sfugge qualcosa di preciso. Prova a rielaborare a voce quello che ha capito, come se stesse spiegando il concetto a qualcuno. L'interfaccia ascolta e, a partire da quanto traspare della sua comprensione, non le restituisce una spiegazione generica del paragrafo ma risponde esattamente a ciò che manca nella sua rielaborazione, correggendo il fraintendimento specifico o aggiungendo l'elemento che non aveva considerato. La spiegazione viene fornita via audio, il che dà a Bianca l'impressione di star avendo una conversazione. Oggi è sola, ma quando l'ambiente circostante non le consente di fare rumore, lo stesso scambio avviene in forma scritta.

Il giorno dopo Bianca legge un romanzo per rilassarsi.

Trova un'espressione che non ha mai sentito: di norma l'avrebbe ignorata, ma l'interfaccia gliene spiega il significato senza doversi spostare dal testo. L'espressione la incuriosisce: chiede se c'è un motivo dietro di essa, e l'interfaccia le racconta l'origine e qualche curiosità. Non è sicura della pronuncia, quindi chiede di sentirla.

Essendo immersa nella lettura in inglese, le viene naturale ricevere le spiegazioni in inglese.

**Mattia - Studiare l'inglese attraverso la lettura** Mattia ha deciso di ritagliarsi un'ora al giorno per esporsi all'inglese fuori dal corso, consapevole che le ore di lezione non bastano. Si trova a casa davanti alla sua scrivania e sceglie di leggere un blog di cinema: un argomento che conosce, che gli permette di seguire il testo senza abbandonarlo, ma che lo mette comunque di fronte a un vocabolario che non padroneggia ancora.

Mentre legge incontra varie parole nuove e l'interfaccia gliele spiega a un livello di inglese a lui consono. Dopo un po' comincia ad accusare la fatica: è già alla terza parola che non conosce e non ha più voglia di una spiegazione articolata — vuole solo sapere cosa significa per poter andare avanti. Chiede direttamente la traduzione. Sa che il sistema salverà le parole incontrate e potrà ritrovarle nel testo, quindi non sente il bisogno di appuntarle come farebbe di solito.

A un certo punto nota che nel testo compare una forma grammaticale che ha appena affrontato a lezione. La riconosce, ma non gli è ancora chiaro quando

vada usata rispetto a un'altra forma con cui la confonde spesso. Stavolta non vuole lasciare il dubbio irrisolto. Chiede all'interfaccia perché viene usata in quel punto del testo e spiega con quale altra forma tende a confonderla. L'interfaccia gli chiarisce la logica e perché in quel contesto è più appropriata dell'alternativa. Mattia chiede altri esempi: l'interfaccia gliene propone alcuni e lo invita a provare a costruire una frase, correggendolo o dandogli un feedback positivo.

Più avanti il blog riporta la trascrizione di un'intervista a un giovane attore americano. Mattia si ritrova di fronte a espressioni di slang americano che non ha mai sentito e di cui non riesce a intuire il significato. L'interfaccia gliene spiega il significato, la pronuncia e i contesti d'uso, ricollegandole a riferimenti che sa essere di suo interesse.

## 4.2 Sintesi dei bisogni e definizione dello scope

Questa sezione presenta i bisogni emersi dall'analisi dei context scenarios e il processo con cui sono stati tradotti nelle priorità e nello scope del prototipo.

### 4.2.1 Bisogni emersi

L'analisi dei context scenarios ha permesso di identificare i bisogni riportati in Tabella 4.1, organizzati per persona e preceduti dai bisogni condivisi tra tutti i profili

| **Categoria** | **Bisogno** |
|---|---|
| Condiviso | Risolvere dubbi nel momento in cui emergono, senza dover ricorrere a strumenti esterni o interrompere il contesto di lettura. |
| Condiviso | Ricevere una spiegazione di un termine o espressione coerente con il significato specifico che assume nel testo che si sta leggendo. |
| Condiviso | Ricevere un supporto proporzionato all'investimento che si è disposti a fare in quel momento, variabile a seconda del contesto e dell'obiettivo della lettura. |
| Condiviso | Ritrovare le spiegazioni dei dubbi risolti durante la lettura direttamente nel contesto in cui sono emersi, senza doverle cercare in uno strumento separato. |
| Laura | Risolvere un dubbio lessicale o concettuale nel momento in cui emerge, senza perdere il filo della lettura. |
| Laura | Avere la certezza di aver compreso correttamente un passaggio prima di procedere, specialmente quando la comprensione è rilevante per il lavoro che si sta svolgendo. |
| Laura | Non perdere le informazioni rilevanti incontrate durante la lettura, per poterle riutilizzare nel proprio lavoro. |
| Bianca | Colmare lacune culturali o storiche che ostacolano la comprensione del testo. |
| Bianca | Verificare la propria comprensione di un passaggio complesso quando ci si rende conto di essersi persi. |
| Bianca | Approfondire un termine o un'espressione che si trova interessante, quando se ne ha voglia, senza che questo venga imposto. |
| Bianca | Consolidare termini nuovi che si trovano interessanti, senza dover gestire manualmente un sistema di appunti. |
| Bianca | Conoscere la pronuncia corretta di termini nuovi incontrati durante la lettura. |
| Mattia | Ricevere spiegazioni calibrate al proprio livello linguistico, che non presuppongano conoscenze non ancora acquisite. |
| Mattia | Capire la logica d'uso di una forma linguistica incontrata nel testo, specialmente quando la si confonde con forme simili già note. |
| Mattia | Potersi esercitare con termini e forme linguistiche incontrati durante la lettura, in un momento a propria scelta. |
| Mattia | Non perdere i termini nuovi incontrati durante la lettura, per poterli ripassare in seguito. |
| Mattia | Ricevere supporto attivo nel consolidamento dei termini nel tempo. |

**Tabella 4.1:** Bisogni emersi dall'analisi dei context scenarios, condivisi e per persona.

### 4.2.2 Definizione delle priorità e scelte di scope

I bisogni elencati nella sezione precedente sono stati derivati dall'analisi dei context scenarios, che immaginano ciascuna persona interagire con un sistema ideale capace di rispondere a tutti i bisogni emersi dallo studio formativo. Prima di tradurli in requisiti, è stato necessario sottoporli a un processo di definizione delle priorità basato su tre criteri: la pertinenza al contesto di lettura, la priorità delle personas e la fattibilità nel prototipo.

Il primo criterio ha portato a escludere dallo scope i bisogni che richiederebbero di accompagnare l'utente al di fuori della sessione di lettura, in primo luogo il consolidamento sistematico del vocabolario attraverso il ripasso e la pratica produttiva estesa. Riconoscere questo limite significa anche riconoscere che lo strumento non può porsi come obiettivo il supporto completo all'acquisizione linguistica. Come emerge dalla letteratura sull'apprendimento lessicale in L2, l'acquisizione richiede sia esposizione contestualizzata di qualità sia incontri ripetuti nel tempo, e quest'ultimo aspetto trascende il perimetro della lettura singola. Lo strumento si concentra quindi sul primo contributo, offrendo al lettore incontri con parole ed espressioni di qualità sufficiente da favorirne un processamento profondo e aumentare le probabilità di ritenzione, senza sostituire un percorso di consolidamento strutturato.

Il secondo criterio ha richiesto di stabilire quale persona imponga i vincoli di design più stringenti da soddisfare simultaneamente alle altre, non quale sia più rappresentata nel campione. Applicato all'intero insieme di bisogni emersi dai context scenarios, questo criterio indicherebbe Mattia, i cui bisogni includono l'unico bisogno di potersi esercitare in una produzione linguistica guidata e il consolidamento nel tempo. Tuttavia il primo criterio di scope, per sua natura, esclude a monte proprio i bisogni che richiedono di accompagnare l'utente fuori dalla sessione di lettura, indipendentemente da quale persona li esprima. Applicato a valle di questo filtro, l'insieme dei bisogni di Mattia rilevanti per lo strumento si riduce a un sottoinsieme di quelli condivisi e di quelli di Bianca, come le spiegazioni contestualmente accurate e calibrate al livello linguistico, senza introdurre un vincolo di design distinto. All'interno dello scope così definito, è Bianca a esprimere il bisogno più difficile da soddisfare senza compromettere gli altri profili, ovvero tenere insieme una risoluzione dei dubbi a bassissimo attrito e un approfondimento facoltativo orientato all'apprendimento, senza che quest'ultimo venga mai imposto. È servendo Bianca che lo strumento serve anche Laura, che si ferma al primo livello di supporto, e Mattia, che utilizza anche il secondo livello con calibrazione linguistica, con aggiustamenti minori. Bianca è pertanto designata come persona primaria. Mattia è una persona secondaria, in quanto condivide il nucleo dei bisogni di Bianca, e lo strumento include aggiustamenti specifici per alcune sue esigenze aggiuntive legate al contesto di lettura, senza che questi compromettano

l'esperienza della persona primaria. Laura è una persona supplementare, i cui bisogni primari, efficienza nella comprensione e continuità con il proprio lavoro, sono in larga parte già servibili da strumenti esistenti. Lo strumento le offre valore, ma non è progettato specificamente per lei.

Il terzo criterio ha portato a rinviare alcune funzionalità possibilmente integrabili al contesto di lettura, in particolare la riproduzione audio e l'input vocale, non per ragioni di scope concettuale, ma per la volontà di validare innanzitutto l'interazione e il design centrali dello strumento, senza introdurre funzionalità aggiuntive che avrebbero a loro volta richiesto una propria verifica. Questa scelta è stata particolarmente rilevante in vista della successiva valutazione dell'effetto dello strumento sulla ritenzione lessicale, discussa più avanti in questa tesi, poiché includere fin da subito più funzionalità avrebbe reso più difficile isolare il contributo di ciascuna. Lo stesso principio guida, più avanti nel processo, anche la scelta di fissare la lingua di interazione. Queste funzionalità sono pertanto riconosciute come rilevanti ma classificate come funzionalità rimandate.

Le priorità e i requisiti definiti in questo capitolo costituiscono il punto di partenza del processo di design descritto nel capitolo successivo, mentre alcuni aspetti dello scope, in particolare riguardo alla modalità di produzione linguistica guidata, sono stati ulteriormente rivisti nel corso della successiva fase di implementazione §6.1.

## 4.3 Formalizzazione dei requisiti

I bisogni e i criteri di scope discussi nella sezione precedente vengono qui tradotti in requisiti funzionali e non funzionali dello strumento.

### 4.3.1 Requisiti funzionali

I requisiti funzionali elencati in Tabella 4.2 discendono per la maggior parte direttamente dai bisogni identificati nella sezione precedente, con il solo meccanismo di realizzazione definito nella fase di design successiva. Fanno eccezione l'Approfondimento contestuale e la modalità di produzione linguistica guidata, la cui stessa formulazione come requisito nasce all'interno del processo di design descritto nel capitolo successivo, piuttosto che da una derivazione diretta dai bisogni.

| ID | Nome |
|---|---|
| **FR1** | **Scelta del Testo** |
| FR1.1 | Supporto alla lettura di testi scelti dall'utente |
| **FR2** | **Selezione e Risposta Immediata** |
| FR2.1 | Rilevamento della selezione testuale |
| FR2.2 | Fornitura di una risposta immediata alla selezione senza interrompere la lettura |
| **FR3** | **Approfondimento Contestuale** |
| FR3.1 | Classificazione automatica della selezione per tipologia linguistica |
| FR3.2 | Possibilità di ridefinire la classificazione assegnata |
| FR3.3 | Accesso a contenuti di approfondimento differenziati per tipologia di selezione |
| FR3.4 | Possibilità di porre domande libere sulla selezione |
| FR3.5 | Gestione di interazioni multiple sulla stessa selezione |
| **FR4**[1] | **Modalità Esercitati** |
| FR4.1 | Modalità di produzione linguistica guidata sulla selezione |
| FR4.2 | Fornitura di feedback sulla risposta prodotta dall'utente |
| **FR5** | **Annotazioni Persistenti** |
| FR5.1 | Identificazione persistente delle selezioni già esplorate nel testo |
| FR5.2 | Accesso alle interazioni salvate a partire dalla selezione nel testo |
| **FR6**[2] | **Preferenze Linguistiche** |
| FR6.1 | Selezione della lingua delle risposte (italiano o inglese) |
| FR6.2 | Persistenza della preferenza linguistica tra sessioni |

[1] Implementato nel prototipo iniziale e successivamente rimosso per conflitto con NFR2; cfr. §6.1.

[2] Rimosso dal prototipo valutato per garantire una condizione sperimentale controllata; cfr. §6.1.

**Tabella 4.2:** Requisiti funzionali del sistema.

### 4.3.2 Requisiti non funzionali

I requisiti non funzionali derivano dai principi enunciati al termine della §3.4.4, dove l'analisi della letteratura SLA e i contributi delle esperte intervistate convergono

verso alcune qualità che il supporto alla lettura dovrebbe possedere per essere pedagogicamente fondato e compatibile con il flusso della lettura.

| **ID** | **Nome** | **Descrizione** |
|---|---|---|
| NFR1 | Accuratezza contestuale | Le risposte del sistema sono ancorate all'uso specifico del termine o dell'espressione nel testo, non a definizioni generiche decontestualizzate. |
| NFR2 | Interruzione minima | Il supporto si integra nella pagina di lettura senza richiedere cambi di contesto applicativo. I processi automatici di lettura non vengono interrotti da intrusioni non richieste. |
| NFR3 | Disponibilità su richiesta | Il sistema si attiva esclusivamente in risposta a un'azione esplicita dell'utente. Non propone interventi, suggerimenti o notifiche non richiesti. |
| NFR4 | Profondità progressiva | Il sistema offre una risposta minima immediata e lascia all'utente la scelta di approfondire. Il livello di dettaglio aumenta solo su richiesta esplicita. |

**Tabella 4.3:** Requisiti non funzionali del sistema.

# Capitolo 5

# Progettazione

Il passaggio dai requisiti alle decisioni di design qui presentate non è stato un processo lineare: si è proceduto per sketch successivi, fino a convergere sui wireframe mostrati in §5.2. I requisiti e le decisioni di design sono il risultato di un processo di coevoluzione avvenuto in modo iterativo. I requisiti non funzionali, in particolare, hanno mantenuto un ruolo di vincolo stabile lungo tutto il processo, orientando e scartando le direzioni di design via via esplorate. La maggior parte dei requisiti funzionali, invece, discendono direttamente dai bisogni identificati in §4: ciò che è stato definito in questa fase è principalmente il meccanismo con cui realizzarli. Fanno eccezione l'Approfondimento contestuale (FR3), in particolare la classificazione automatica delle selezioni, e la modalità Esercitati (FR4): entrambi sono nati all'interno del processo di design qui descritto, più che da una derivazione diretta dai dati dello studio formativo. Per questo motivo, la classificazione in categorie presentata nelle prossime sezioni non va letta come un risultato emerso direttamente dai dati, ma come una sintesi di design che a quei dati resta ancorata.

## 5.1 Decisioni di design

Le decisioni presentate in questa sezione condividono un principio comune: minimizzare il costo di accesso a una spiegazione contestualmente accurata, rendendo la profondità del supporto fornito proporzionale all'investimento che il lettore è disposto a fare in quel momento, senza mai obbligarlo a uscire dal testo.

Le sottosezioni seguenti sviluppano questo principio lungo le dimensioni progettuali che hanno richiesto una decisione esplicita: la collocazione spaziale dell'interazione, la lingua delle risposte, la struttura a livelli di supporto, il riconoscimento automatico del tipo di difficoltà e la persistenza delle interazioni nel tempo.

### 5.1.1 Approccio inline

Inizialmente era stata contemplata la possibilità di offrire gli approfondimenti relativi al testo all'interno di un pannello laterale. Questa opzione è stata scartata perché avrebbe rischiato di forzare il lettore a distogliere l'attenzione dal corpo del testo. Si è deciso quindi che ogni interazione dovesse avvenire all'interno del testo stesso, in modo da rendere naturale sia il ritorno alla lettura sia il confronto tra la spiegazione ottenuta e la porzione di testo a cui si riferisce.

Da questa scelta discende un vincolo più generale, che governa ogni elemento di supporto mostrato durante la lettura: il supporto deve occupare il minor spazio possibile e non deve mai rendere inaccessibili le parti di testo che precedono o seguono la porzione in cui è emersa la difficoltà, così da permettere al lettore di interpretarlo alla luce del testo circostante. Lo stesso principio di subordinazione spaziale del supporto al testo si traduce anche nella larghezza della colonna di lettura: questa occupa la porzione centrale e dominante dello schermo, e ogni elemento di interfaccia resta vincolato alla stessa larghezza, senza mai eccederla lateralmente. Questo secondo vincolo traduce spazialmente NFR2: se il supporto disponesse di un'area propria indipendente dal testo, per quanto verticalmente inline, il testo non sarebbe più l'elemento centrale dell'esperienza. Vincolarne ogni elemento alla larghezza della colonna rende esplicito, anche visivamente, che l'interazione con il supporto resta subordinata alla lettura e non la sostituisce.

### 5.1.2 Lingua nello strumento

Inizialmente si era deciso di lasciare all'utente la scelta di ricevere il supporto dell'IA in italiano o in inglese. La letteratura consultata non permette di stabilire una preferenza netta tra le due lingue di risposta: l'evidenza tende a favorire la L1 per la ritenzione lessicale, ma il vantaggio è moderato dal livello del lettore, e le spiegazioni in L2 offrono comunque un'esposizione aggiuntiva alla lingua che la traduzione non veicola (§3.7.2). Lasciare la scelta all'utente evitava di dover arbitrare tra queste evidenze in tensione tra loro, permettendo a ciascun lettore di regolarsi secondo la propria preferenza.

La scelta di lasciare la lingua configurabile rispondeva inoltre a un'esigenza osservata trasversalmente nello studio formativo: la traduzione risultava il bisogno più diffuso anche tra i partecipanti di livello avanzato, per i quali la comprensione di una spiegazione in inglese non costituiva una difficoltà. La preferenza per la traduzione non era quindi legata al solo livello di competenza linguistica, ma alla rapidità con cui permette di risolvere un dubbio senza interrompere la lettura; lasciare la scelta all'utente permetteva di rispettare questa preferenza senza escludere, per chi lo desiderava, la possibilità di ricevere risposte in L2.

La fase di design ha previsto quindi un elemento dell'interfaccia che permettesse all'utente di selezionare questa preferenza. Nonostante ciò, nel prototipo valutato

la lingua di interazione con l'IA è stata infine fissata sull'italiano(§6.1). Restano tuttavia sempre in inglese, all'interno delle risposte, gli esempi e i richiami diretti ai termini o alle espressioni oggetto della spiegazione.

### 5.1.3 Livelli di approfondimento

La tensione centrale che questa decisione di design cerca di risolvere è quella tra i requisiti non funzionali di interruzione minima e disponibilità su richiesta (NFR2, NFR3) da un lato, e dall'altro l'Involvement Load Hypothesis (§2.2), secondo cui una maggiore elaborazione cognitiva favorisce la ritenzione lessicale. Coerentemente con quanto osservato nello studio formativo, dove l'apprendimento durante la lettura emerge come obiettivo secondario ma presente, si è deciso di distribuire il supporto su più livelli. Ciascun livello richiede gradualmente maggiore attenzione da parte dell'utente e interrompe quindi in misura crescente il flusso della lettura, risultando però anche più efficace ai fini dell'apprendimento e dell'elaborazione della lingua.

All'utente viene proposto di default sempre e solo il livello di interruzione minima, sufficiente a risolvere il dubbio occasionale più frequente dovuto all'incomprensione e a permettere una lettura scorrevole. Il contenuto di questo primo livello dipende dalla lingua di risposta impostata: una traduzione se italiano, una spiegazione sintetica o un sinonimo contestualizzato se inglese. Da questo aiuto è comunque lasciata all'utente la possibilità di approfondire, ricevendo un supporto via via più consistente.

Il secondo livello rende disponibile la possibilità di ricevere approfondimenti volti a favorire un incontro lessicale di maggiore qualità: interazioni con il testo che non si limitano alla traduzione, ma permettono di conoscere il termine o l'espressione incontrata sotto nuove sfumature.

L'ultimo livello di questa gerarchia, denominato "Esercitati", prevede una modalità orientata ad attività produttive, in cui l'utente può mettere in pratica le nuove forme lessicali incontrate nel momento stesso in cui sorge la curiosità durante la lettura, cimentandosi in un'attività che richiede una partecipazione attiva, arricchita da un feedback in tempo reale fornito dall'intelligenza artificiale. Questa modalità è stata implementata nel prototipo iniziale e successivamente rimossa; le motivazioni sono discusse in §6.1.

### 5.1.4 Classificazione automatica

A differenza delle altre decisioni di design presentate in questo capitolo, la classificazione automatica non discende direttamente da un requisito definito nel Capitolo 4, ma è il risultato di una sintesi di design elaborata a partire dall'osservazione delle difficoltà più comuni riscontrate nello studio formativo (§3.4.1).

Ogni tipo di difficoltà richiede un supporto di natura diversa: la spiegazione utile per una forma verbale non coincide con quella utile per un riferimento culturale. Per rispettare il principio dell'interruzione minima (NFR2) offrendo al tempo stesso un supporto contestualmente rilevante e non generico (NFR1), si è deciso che il sistema riconoscesse automaticamente la tipologia della porzione di testo selezionata, tentando così di anticipare il bisogno dell'utente.

Le categorie di difficoltà emerse dallo studio formativo hanno costituito la base di questa classificazione, poi tradotta in cinque categorie operative: vocabolario, espressione idiomatica, forma verbale o fraseale, riferimento culturale o concettuale, frase o passaggio confuso. Il sistema classifica ogni selezione in una di queste categorie e adatta il supporto messo a disposizione dell'utente di conseguenza.

Il processo di classificazione è interamente interno al sistema: l'utente seleziona il testo senza dover indicare a quale categoria appartiene la propria difficoltà. Chiedere all'utente di categorizzare esplicitamente ciò che non ha capito introdurrebbe un passaggio cognitivo ulteriore prima di ricevere aiuto, e presupporrebbe inoltre che l'utente sappia già distinguere, ad esempio, un riferimento culturale da un'espressione idiomatica: una distinzione non sempre ovvia nel momento stesso dell'incomprensione.

### 5.1.5 Persistenza delle interazioni

Questa decisione di design deriva da un bisogno condiviso da tutte le personas: poter ritrovare, in un momento successivo alla lettura, le informazioni rinvenute durante le interazioni con il testo. Si è deciso quindi di prevedere un sistema di annotazione che permette all'utente di ritrovare le porzioni di testo interrogate, insieme alle informazioni rinvenute e ritenute rilevanti.

Una seconda decisione riguarda la continuità delle singole interazioni: se l'utente reinterroga una porzione di testo già esplorata, l'interazione non riparte da zero, ma riprende da dove era stata lasciata. Questo evita che una stessa difficoltà, se rivisitata, richieda di ripetere l'intero percorso di approfondimento già intrapreso.

## 5.2 Wireframes

I wireframe presentati di seguito rappresentano il design delle interazioni verso cui convergono le decisioni appena discusse, così come approvato al termine della fase di progettazione. Non sono vincolati a una piattaforma specifica: l'obiettivo di questa fase era definire il comportamento dell'interazione a prescindere dal dispositivo, rimandando la scelta della piattaforma di implementazione alla fase successiva (§5.3). Sono stati realizzati con Balsamiq, uno strumento dedicato alla creazione di wireframe per interfacce utente.

Questi wireframe costituiscono il punto di partenza della successiva fase di implementazione, condotta per iterazioni successive che hanno portato, sulla base delle necessità e dei risultati riscontrati nel processo, al prototipo descritto nel Capitolo 6 e successivamente utilizzato nello studio di valutazione discusso nel Capitolo 7. Ogni elemento qui presentato e in seguito rimosso o modificato è segnalato come tale nel testo che accompagna la figura corrispondente; le motivazioni dello scostamento sono discusse in §6.1.

### 5.2.1 Interazione minima




**Figura 5.1:** Stato iniziale dell'interfaccia di lettura.

Nello stato iniziale, prima di qualsiasi selezione, la pagina mostra il testo da leggere (Figura 5.1). In cima alla schermata è presente uno switch che permette la selezione della lingua di risposta; questa funzionalità, prevista in fase di design, è stata rimossa durante l'implementazione (§6.1).

Alla risposta immediata, generata in reazione alla selezione di una porzione di testo, corrisponde un elemento minimale contenente la traduzione del termine o dell'espressione selezionata se il sistema è impostato su italiano, o una spiegazione semplificata in inglese se impostato su tale lingua(Figura 5.2). Poiché la fase di design si è concentrata principalmente sulla definizione delle interazioni piuttosto che sul confronto sistematico tra le due varianti linguistiche, i wireframe presentati mostrano unicamente la configurazione in italiano. L'elemento appare in un popover che si sovrappone localmente al testo circostante, posizionato sotto la selezione o, se lo spazio sottostante non è sufficiente, sopra di essa. In questo caso si accetta di oscurare parzialmente il testo, date le dimensioni contenute del popover, pensato per non richiedere alcuna riorganizzazione del layout in un'interazione che l'utente può chiudere immediatamente per continuare la lettura.

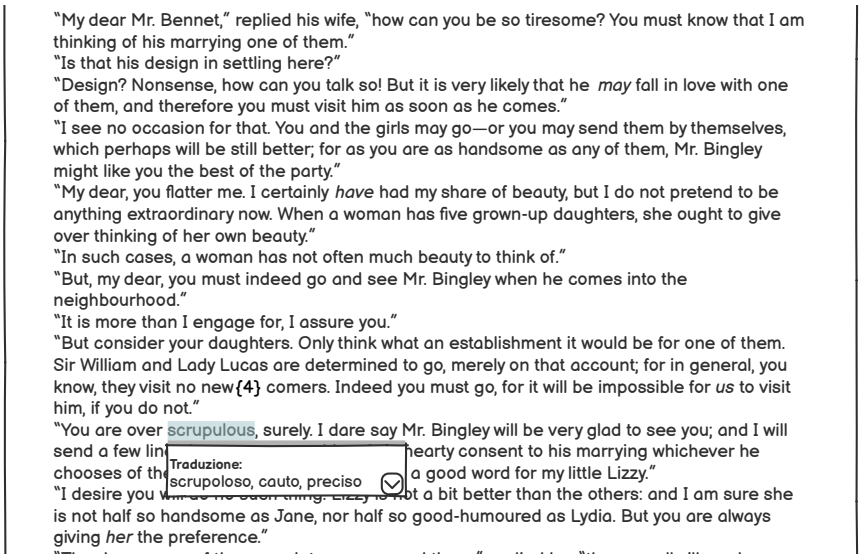



**Figura 5.2:** Selezione di una parola e risposta immediata.

All'interno di questo elemento è presente un bottone per accedere al livello successivo di approfondimento.

### 5.2.2 Pannello inline

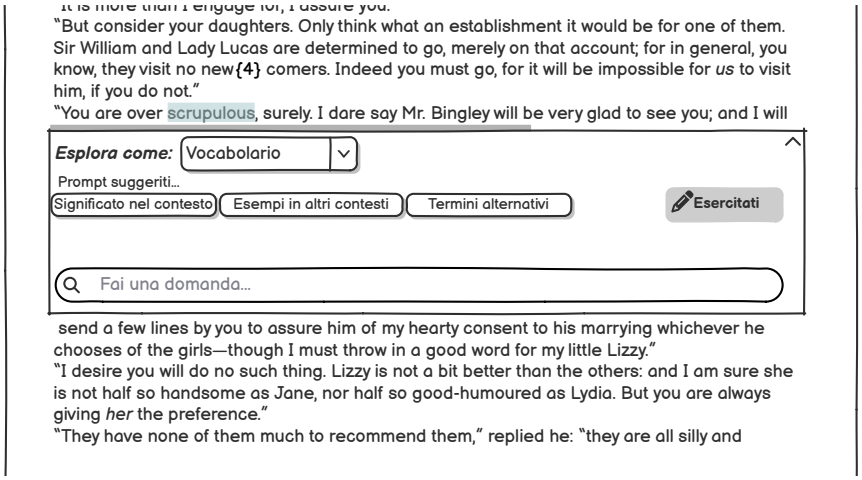



**Figura 5.3:** Pannello inline espanso.

Passando al livello di approfondimento successivo, l'utente sceglie di aprire il pannello di approfondimento inline (Figura 5.3). In fase di design, questo livello era denominato modalità "Comprensione", in contrapposizione alla modalità "Esercitati" (o "Produzione") discussa più avanti. Questa distinzione nominale non è sopravvissuta all'implementazione, insieme alla modalità Esercitati stessa (§6.1).

A differenza dell'elemento minimale del primo livello, il pannello non si sovrappone al testo ma si inserisce nel flusso del documento, spingendo il contenuto successivo verso il basso. Questa scelta risponde alla stessa esigenza di minimo ingombro discussa in §5.1.1, applicata però a un elemento che, dovendo restituire una quantità maggiore di informazioni, non potrebbe più sovrapporsi al testo senza oscurarne porzioni rilevanti: spingendo il contenuto anziché coprirlo, il pannello si premura di lasciare comunque l'intero testo accessibile, pur occupando più spazio sullo schermo. Durante tutta l'interazione, la porzione di testo interessata dalla selezione rimane evidenziata in modo chiaro, e il pannello, come l'elemento minimale del primo livello, resta ancorato ad essa.

In testa al pannello viene mostrato il risultato della classificazione del testo evidenziato, ricondotto a una delle cinque categorie descritte in §5.1.4. A seconda della categoria, vengono proposti come approfondimenti disponibili dei chip cliccabili, ciascuno associato a un prompt predefinito inviato all'IA per ricevere una risposta coerente con la richiesta.

In fondo al pannello è disponibile una barra di ricerca che l'utente può usare per porre liberamente domande riguardanti il testo selezionato. In questo modo si evita che l'utente sia limitato ai soli suggerimenti proposti, potendo formulare

autonomamente un quesito qualora un suo dubbio specifico non venga anticipato da uno degli approfondimenti disponibili.

Il bottone a destra dei chip permette di passare alla modalità Esercitati.

### 5.2.3 Selezione di un prompt

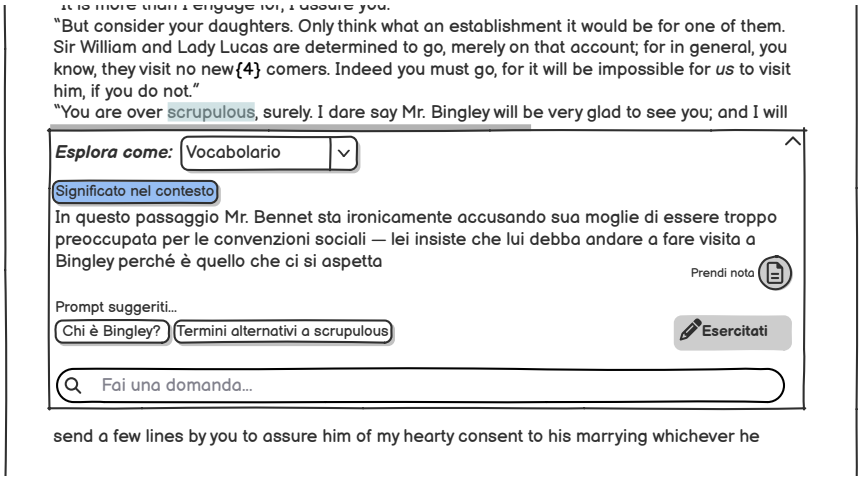
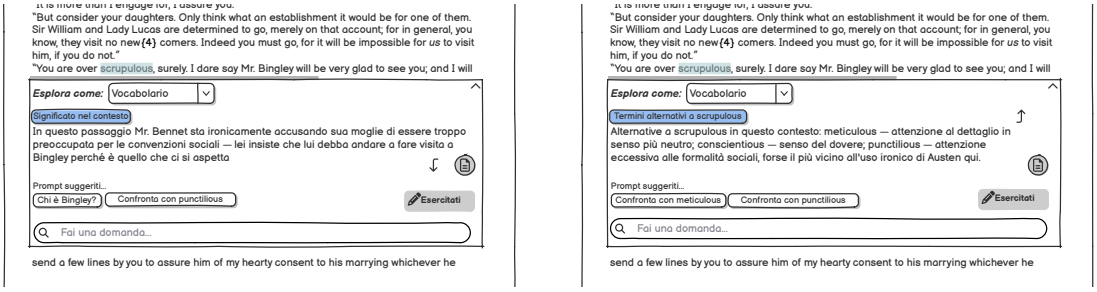



**Figura 5.4:** Prompt selezionato

Cliccando su uno dei chip contenente un approfondimento suggerito, viene fatto spazio per la risposta restituita dall'IA (Figura 5.4). In fase di design, gli altri suggerimenti scomparivano dopo la selezione; in implementazione si è invece scelto di mantenerli sempre disponibili (§6.3.3).

Oltre alla risposta, l'IA genera in questo livello di approfondimento un paio di domande opzionali che l'utente può scegliere di inviare, formulate per seguire la direzione presa dalle domande già poste dall'utente, nel tentativo di anticipare ulteriori possibili dubbi.

Una volta ricevuta più di una risposta sulla stessa selezione, è possibile navigare tra le risposte ottenute, tramite i controlli mostrati nei wireframe come frecce dirette verso l'alto o il basso (Figura 5.5).




**(a)** Possibilità di navigare a risposta successiva

**(b)** Possibilità di navigare a risposta precedente

**Figura 5.5:** Navigazione tra risposte ricevute

Un pulsante presente accanto a ciascuna risposta permette di salvarla nella nota associata alla porzione di testo interessata (Figura 5.6).

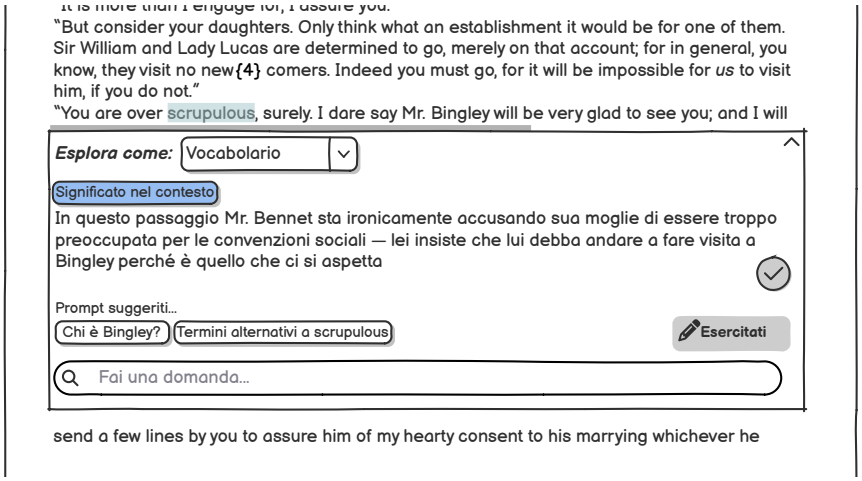



**Figura 5.6:** Risposta salvata nella nota

### 5.2.4 Classificazione automatica e manuale

La classificazione automatica delle porzioni di testo è affidata sempre a un modello linguistico, che può quindi commettere errori. Per questo, coerentemente con quanto anticipato in §5.1.4, è lasciata all'utente la possibilità di correggere la classificazione assegnata dal sistema (Figura 5.7).

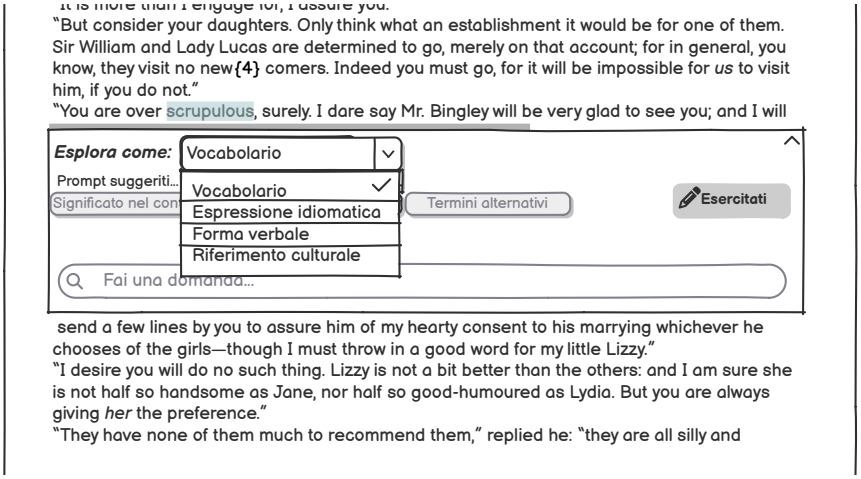
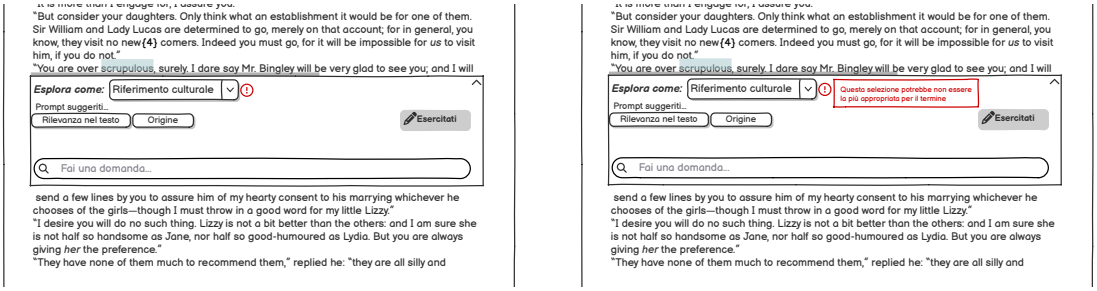



**Figura 5.7:** Apertura menù classificazione

Cliccando sul controllo associato alla categoria assegnata dal sistema, si apre un menu a tendina in cui appaiono le varie categorie possibili, ordinate dall'IA in base alla probabilità che la porzione di testo selezionata vi appartenga. L'utente può selezionare liberamente una categoria diversa, con un conseguente cambio dei prompt di approfondimento suggeriti.




**(a)** Simbolo avviso di categoria sbagliata **(b)** Apertura avviso di categoria sbagliata

**Figura 5.8:** Categoria possibilmente sbagliata

Nel caso l'utente selezioni una categoria che, secondo la classificazione del modello, risulta poco compatibile con il testo selezionato, un segnale di avviso appare accanto alla categoria scelta; al passaggio del mouse, un messaggio a comparsa ne spiega la possibile incompatibilità (Figura 5.8). Se l'utente procede comunque selezionando un prompt, l'IA risponde in ogni caso basandosi sulla categoria scelta dall'utente: in alcuni casi il prompt risulta comunque adeguato al testo nonostante l'incongruenza segnalata, mentre in altri l'IA stessa evidenzia

nella risposta la possibile incompatibilità tra la categoria selezionata e il contenuto della porzione di testo.

### 5.2.5 Modalità Esercitati

La modalità Esercitati appare in una forma analoga a quella rivolta alla comprensione, con alcune differenze (Figura 5.9).

I prompt suggeriti tra cui l'utente può scegliere sono formulati all'imperativo, per segnalare che richiedono una produzione attiva da parte dell'utente. Il campo di input, in questo caso, serve a inviare i propri tentativi e a chiedere chiarimenti su eventuali feedback ricevuti. Un bottone a destra permette di tornare alla modalità di comprensione.

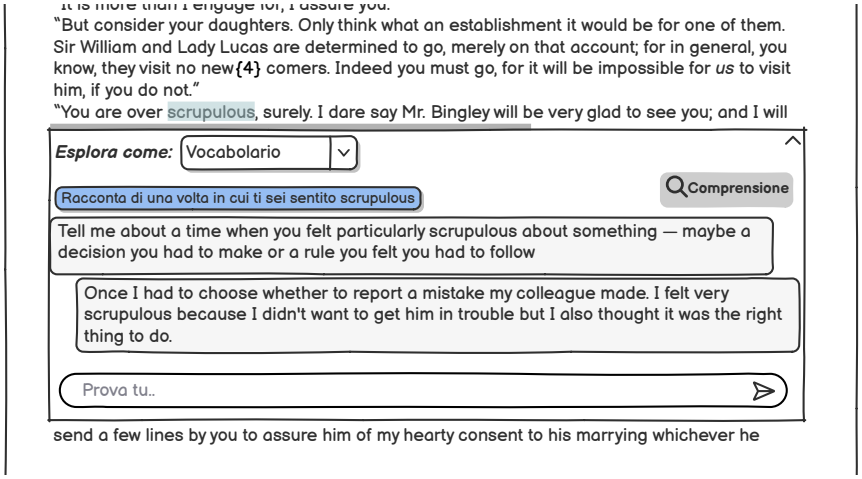



**Figura 5.9:** Modalità "Esercitati"

Cliccando su uno dei prompt, l'IA genera un incipit che compare nel pannello come avvio dell'attività, ad esempio "Tell me about a time when you felt scrupulous about something." L'incipit è sempre formulato in inglese, indipendentemente dalla lingua impostata nel toggle: trattandosi di un'attività di produzione linguistica, ha senso che avvenga nella lingua target, mentre il toggle regola unicamente la lingua del feedback ricevuto. L'utente risponde tramite il campo di input, sempre in inglese; la risposta compare sotto l'incipit in uno scambio in stile chat.

L'IA fornisce un feedback su due dimensioni: contenutistica, relativa alla coerenza del contesto d'uso del termine, e linguistica, relativa a eventuali correzioni grammaticali o suggerimenti di miglioramento. Il feedback è fornito nella lingua impostata nel toggle. Accanto ad esso è presente un bottone per salvare l'interazione nella nota, che ne conserva incipit, risposta dell'utente e feedback in forma compatta.

Questa modalità è stata implementata nel prototipo iniziale e successivamente rimossa in fase di implementazione; le motivazioni sono discusse in §6.1.

### 5.2.6 Note persistenti

Quando l'utente chiude il pannello, cliccando fuori da esso o su un bottone di chiusura apposito, la porzione di testo con cui ha interagito rimane sottolineata, a segnalare la presenza di una nota associata ad essa (§5.1.5). La nota si crea indipendentemente dal numero di livelli esplorati: contiene sempre la risposta all'interazione di primo livello, ovvero la traduzione o la spiegazione semplificata a seconda della lingua impostata, ed eventuali interazioni avvenute all'interno del pannello di approfondimento, se attivamente salvate dall'utente.

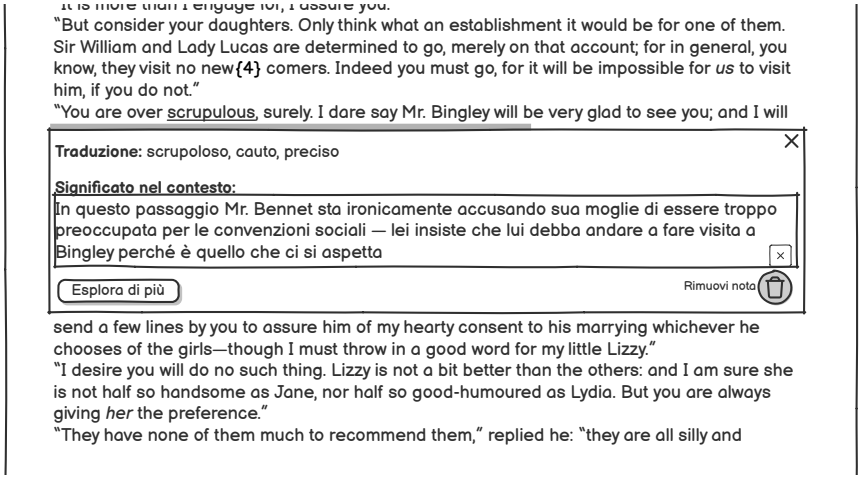



**Figura 5.10:** Nota inline

Cliccando sulla porzione di testo sottolineata si accede alla nota, che nei wireframe qui presentati appare sempre inline e ancorata al testo a cui è associata (Figura 5.10). In fase di implementazione questa collocazione è stata rivista; le motivazioni sono discusse in §6.3.3. All'interno della nota, l'utente può eliminare le singole risposte salvate, cancellare la nota nel suo insieme, o tornare al pannello di approfondimento tramite il bottone "Esplora di più".

## 5.3 Architettura del sistema

La piattaforma scelta per il prototipo è un'applicazione web per desktop. Tra i due dispositivi più usati dal campione per la lettura in inglese, computer (n=72 su 89) e smartphone (n=78), le interviste indicano una differenza qualitativa nella modalità

di lettura associata a ciascuno. Lo smartphone è infatti solitamente utilizzato per la lettura di contenuti leggeri e scorrevoli, mentre il computer per una lettura più impegnativa e attenta. È quindi generalmente associato a una maggiore tolleranza all'attrito, mentre lo smartphone è associato quasi universalmente a una lettura di bassa profondità e a una bassa tolleranza per qualsiasi interruzione. Poiché lo strumento è pensato per un tipo di lettura strutturata e per un supporto che, per quanto minimizzato, introduce comunque un certo grado di interazione, il computer risulta il contesto d'uso più coerente. La scelta non esclude un adattamento a piattaforme diverse in sviluppi futuri dello strumento.

A livello concettuale, il sistema è stato pensato secondo un'architettura client-server standard, articolata su tre componenti principali. Il lato client gestisce l'interfaccia di lettura, la selezione del testo e la visualizzazione delle risposte ai diversi livelli di approfondimento, come anche la presentazione delle annotazioni persistenti. Il lato server orchestra le richieste verso il modello linguistico, gestendo la classificazione, la generazione delle risposte e il mantenimento del contesto per le interazioni multi-turno, oltre al salvataggio e recupero delle annotazioni. Due risorse esterne completano il sistema: un servizio di generazione basato su modello linguistico, responsabile di classificazione, spiegazioni contestuali, esercizi e feedback, e un database per la persistenza delle annotazioni e della cronologia delle interazioni tra sessioni diverse.

La Figura 5.11 sintetizza questa architettura concettuale, che ha guidato la successiva fase di implementazione descritta nel Capitolo 6.

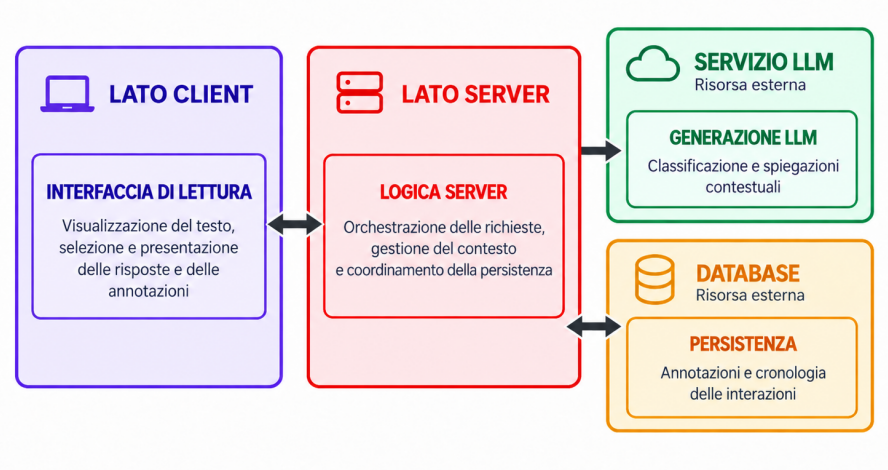



**Figura 5.11:** Architettura concettuale del sistema

# Capitolo 6

# Implementazione

All'interno di questo capitolo viene mostrato come le decisioni prese durante la fase di progettazione sono state messe in pratica e sviluppate attraverso un processo iterativo, che ha portato al prototipo nella sua forma finale, ovvero quella utilizzata nella successiva fase di validazione. Il sistema realizzato è stato denominato "ReadLens".

In primo luogo vengono esposte le motivazioni dietro ai cambiamenti più sostanziali effettuati durante l'implementazione, che hanno portato il prototipo a discostarsi in parte dal design definito nella fase di progettazione. Si entra poi nel dettaglio di come le componenti del sistema, già introdotte a livello concettuale in §5.2, sono state effettivamente implementate, quali tecnologie sono state utilizzate e come sono state strutturate internamente.

## 6.1 Scelte progettuali in corso d'opera

Due sono stati i cambiamenti più sostanziali effettuati durante l'implementazione dello strumento, tra un'iterazione e l'altra.

Il primo riguarda la rimozione della modalità Esercitati (FR4), presentata in §5.1.3 e §5.2.5 come ultimo livello della gerarchia di approfondimento volta a favorire occasioni di apprendimento gradualmente più efficaci. Nonostante il potenziale di questa modalità, una volta implementata e testata si è ritenuto che l'interazione da essa prevista costituisse un'interruzione troppo rilevante per la lettura del testo, in conflitto con l'obiettivo di interruzione minima (NFR2). Si è deciso quindi di limitarsi a quella che in fase di design era stata denominata modalità Comprensione, rendendola l'unica modalità disponibile all'interno del pannello inline, rinunciando consapevolmente a parte del contributo potenziale all'apprendimento del lettore a favore dell'usabilità e della naturalezza di fruizione dello strumento. La modalità

Esercitati è stata rimandata a lavori futuri, nella prospettiva di poterla integrare in una forma diversa, ad esempio come attività successiva alla lettura.

Il secondo cambiamento sostanziale ha riguardato la rimozione della possibilità di impostare la lingua di interazione con lo strumento (FR6). Dopo aver implementato tale funzionalità, che rendeva possibile il cambio di lingua tramite uno switch in cima alla pagina, si è deciso di renderla inaccessibile e fissare la lingua all'italiano, lasciando sempre in lingua originale eventuali esempi o richiami a termini ed espressioni inglesi. Questa decisione è stata presa principalmente nell'ottica di ottenere una condizione sperimentale più pulita. Considerata la volontà di testare il contributo dello strumento alla comprensione e all'apprendimento, lasciare all'utente la possibilità di ricevere supporto sia in italiano sia in inglese avrebbe introdotto un'ulteriore variabile nel confronto, il cui contributo specifico sarebbe stato difficile da isolare. La scelta di fissare la lingua sull'italiano, piuttosto che sull'inglese, è stata inoltre coerente con l'evidenza discussa in §2.4.1, che tende debolmente a favorire la L1 per la ritenzione lessicale nei contesti di glossing, oltre che con i bisogni espressi dalle personas verso un supporto immediato e a basso attrito. Anche questa funzionalità è stata perciò rimandata a sviluppi futuri dell'applicazione (§8.1).

Ulteriori modifiche, di natura più circoscritta e legate a scelte di dettaglio nell'interfaccia, sono discusse contestualmente alla loro presentazione in §6.4.

## 6.2 Tecnologie

Come anticipato in §5.3, il prototipo è basato su un'architettura client-server, che separa quindi la gestione dell'interfaccia utente (frontend) dalla logica applicativa e dall'orchestrazione delle richieste verso il modello linguistico (backend).

È stato sviluppato interamente in TypeScript, sia lato backend sia lato frontend, scelto per la sicurezza di tipo che previene un gran numero di errori già in fase di compilazione, oltre che per la sua ampia adozione nell'ecosistema React. Il versionamento del codice, per l'intero progetto, è stato gestito con Git.

Il frontend è realizzato in React, scelto per il modello a componenti che si presta bene a un'interfaccia costruita attorno a stati di interazione multipli (popover, pannello, note), con Vite come strumento di build e sviluppo, apprezzato per i tempi di ricarica rapidi durante lo sviluppo iterativo dell'interfaccia. Lo stile è definito con Tailwind CSS, integrato con i componenti della libreria shadcn, che ha permesso di velocizzare la realizzazione di elementi di interfaccia comuni (menu a tendina, pulsanti, campi di input) senza doverli costruire da zero.

Il backend è costruito su Node.js, con Express come framework per la gestione del server e delle rotte API. La build del codice TypeScript è affidata a tsc. L'integrazione con il modello linguistico avviene tramite l'SDK ufficiale di OpenAI,

utilizzando il modello GPT-4o-mini, scelto per il buon compromesso tra qualità delle risposte, velocità e costo per richiesta, rilevante data la frequenza delle interazioni previste durante una sessione di lettura. Le librerie cors e dotenv gestiscono rispettivamente le policy di condivisione delle risorse tra origini diverse e la configurazione delle variabili d'ambiente. La persistenza dei dati si appoggia su un database relazionale PostgreSQL, gestito tramite l'ORM Prisma, che permette di definire lo schema dei dati e interrogare il database senza scrivere query SQL manualmente. Le dipendenze sono gestite con npm.

La comunicazione tra frontend e backend avviene tramite richieste REST su HTTP, con payload in formato JSON, effettuate tramite le API fetch native del browser.

## 6.3 Frontend

Questa sezione descrive come le decisioni di design e i wireframe presentati nel Capitolo 5 sono stati tradotti nell'interfaccia di ReadLens. Viene innanzitutto descritta l'organizzazione generale del codice frontend, seguita da una panoramica delle schermate principali, dei componenti che le compongono e della gestione dello stato dell'applicazione.

### 6.3.1 Architettura e organizzazione

Il frontend è organizzato attorno a tre rotte principali, gestite con React Router: la libreria dei documenti, l'importazione di un nuovo testo e la vista di lettura, dove avviene la quasi totalità delle interazioni descritte in §5.2.

Lo stato dell'applicazione vive interamente in hook locali, istanziati dai componenti che li utilizzano, senza un livello di stato condiviso a livello globale. Questa scelta è coerente con la natura dell'applicazione, pensata per un singolo utente e priva di dati che debbano essere sincronizzati tra rotte diverse.

Il recupero dei dati non è concentrato in un livello a sé, ma collocato dove serve, con tempistiche diverse a seconda del caso. Alcuni dati infatti vengono richiesti immediatamente all'apertura di una pagina, ad esempio l'elenco dei documenti o le annotazioni già presenti in un testo, mentre altri vengono richiesti solo in risposta a un'interazione specifica dell'utente, come una selezione di testo o l'apertura del pannello di approfondimento.

La logica applicativa è in gran parte incapsulata in hook dedicati, ciascuno responsabile di una singola competenza verticale, come la gestione della selezione testuale, la classificazione, la generazione delle risposte o la persistenza delle annotazioni. La composizione di più hook all'interno di un componente svolge, limitatamente al proprio sottoalbero, un ruolo analogo a quello di uno stato condiviso, pur senza fornire l'accessibilità cross-componente che caratterizzerebbe

uno store applicativo centralizzato. Il componente che orchestra la vista di lettura, in particolare, non si limita a comporre questi hook ma contiene anche la logica di rendering necessaria a gestire correttamente porzioni di testo con annotazioni sovrapposte o a cavallo tra più paragrafi.

### 6.3.2 Linee guida per l'interazione con l'IA

Alcune decisioni di design del frontend sono state guidate dalle linee guida per l'interazione uomo-IA proposte da Amershi et al. (2019) [27], un insieme di 18 principi rivolti alla progettazione di sistemi che integrano intelligenza artificiale nell'esperienza utente. La Tabella 6.1 elenca le linee guida richiamate in questo capitolo; nel seguito, le decisioni riconducibili a una linea guida specifica la citano esplicitamente tramite la propria sigla.

| **Sigla** | **Nome** | **Principio** |
|---|---|---|
| G2 | Make clear how well the system can do what it can do | Comunicare l'affidabilità del sistema, ad esempio segnalando la natura generata di un contenuto |
| G4 | Show contextually relevant information | Mostrare informazioni pertinenti al contesto specifico dell'utente |
| G7 | Support efficient invocation | Rendere facile richiedere il supporto dell'IA quando serve |
| G8 | Support efficient dismissal | Rendere facile ignorare o chiudere un supporto non desiderato |
| G9 | Support efficient correction | Rendere facile correggere il sistema quando sbaglia |
| G15 | Encourage granular feedback | Permettere all'utente di esprimere un giudizio sulle singole risposte ricevute |
| G16 | Convey the consequences of user actions | Comunicare come le azioni dell'utente influenzeranno il comportamento futuro del sistema |

**Tabella 6.1:** Linee guida di Amershi et al. (2019) richiamate nel capitolo.

L'intero impianto dell'interazione, con la sua gerarchia di livelli attivabili solo su richiesta (§5.1.3), è coerente in generale con i principi di invocazione e interruzione efficienti (G7,invocazione efficiente; G8,chiusura efficiente) e con la presentazione di

informazioni pertinenti al contesto specifico dell'utente (G4,informazioni contestualmente rilevanti). Altre linee guida hanno invece orientato decisioni più puntuali, discusse nel dettaglio in seguito.

### 6.3.3 Stati dell'interfaccia

Questa sezione presenta gli stati principali dell'interfaccia nella loro versione finale, usata nella successiva fase di validazione. Eventuali discostamenti dal design iniziale presentato nei wireframes vengono segnalati e motivati. Il testo utilizzato come esempio negli screenshot che seguono è tratto da una recensione cinematografica pubblicata online, usata unicamente a scopo illustrativo[1].

#### La libreria

All'apertura dell'applicazione, l'utente accede alla propria libreria personale (Figura 6.1), dove i testi importati sono mostrati in ordine di ultimo accesso. Se la libreria non contiene ancora alcun documento, un messaggio invita l'utente a importarne uno.

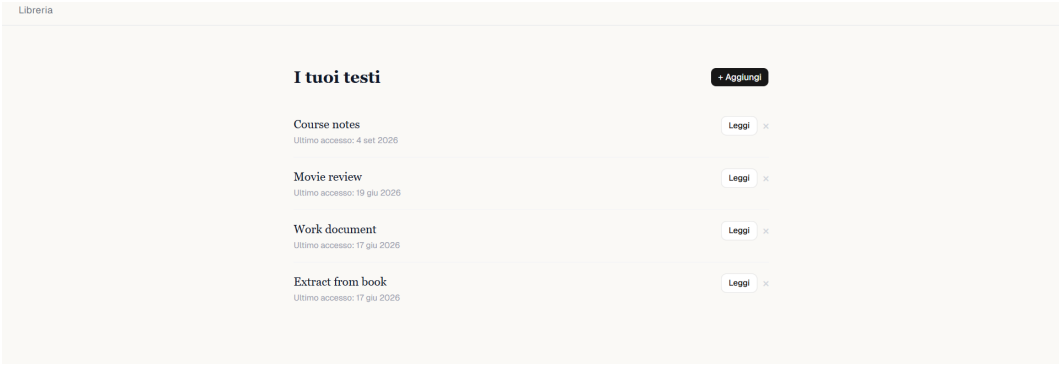



**Figura 6.1:** Libreria con alcuni testi già importati.

Nella pagina dedicata all'importazione, l'utente può aggiungere un nuovo testo indicandone il titolo e fornendo il contenuto tramite un file .txt o incollandolo direttamente nell'area di testo prevista (Figura 6.2). Questa scelta è stata consapevole: non essendo ancora stato validata l'interazione centrale dello strumento da parte degli utenti, non si è ritenuto prioritario investire tempo nel supporto a formati di importazione più articolati, preferendo destinarlo al perfezionamento dell'interazione di lettura. Un supporto più ampio ai formati di importazione, così

[1] Hamnet|Review, thefilm.blog, disponibile su `[https://thefilm.blog/2026/01/22/hamnet-review/]` (consultato il [15/05/2026]).

come una possibile estensione dello strumento verso un'estensione browser, restano considerati per sviluppi futuri.

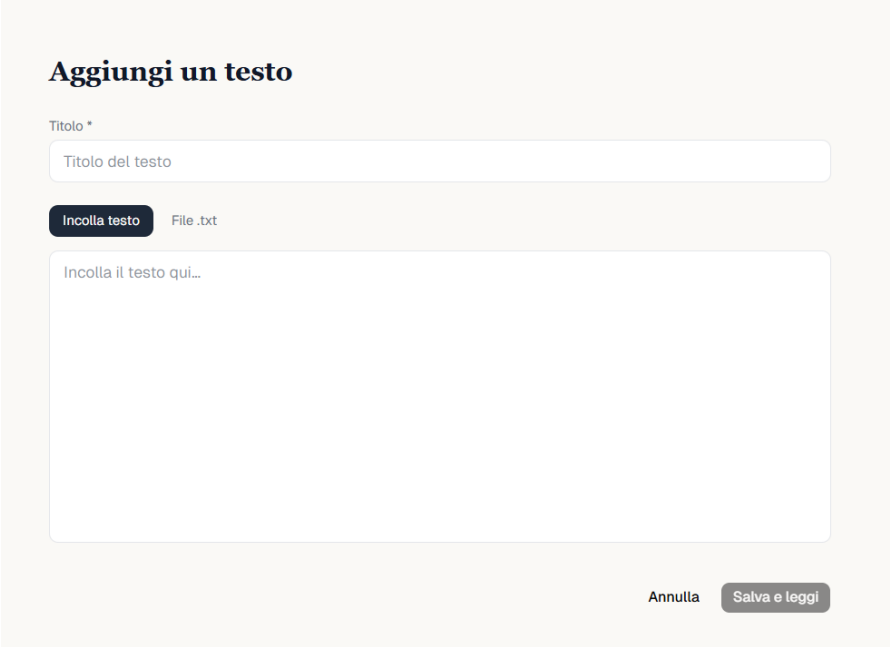



**Figura 6.2:** Form per l'aggiunta di un nuovo testo.

### Schermata di lettura

Quando l'utente sceglie di leggere uno dei testi importati, tramite il pulsante "Leggi" nella libreria, il testo si apre nella pagina di lettura, occupando la porzione centrale dello schermo (Figura 6.3). Ciascuna riga contiene circa 65 caratteri, una misura comunemente indicata nella pratica di design delle interfacce come favorevole alla leggibilità di un testo continuo, in quanto evita sia righe troppo corte, che spezzano il ritmo di lettura, sia righe troppo lunghe, che rendono difficile per l'occhio non perdere il filo da una riga all'altra.

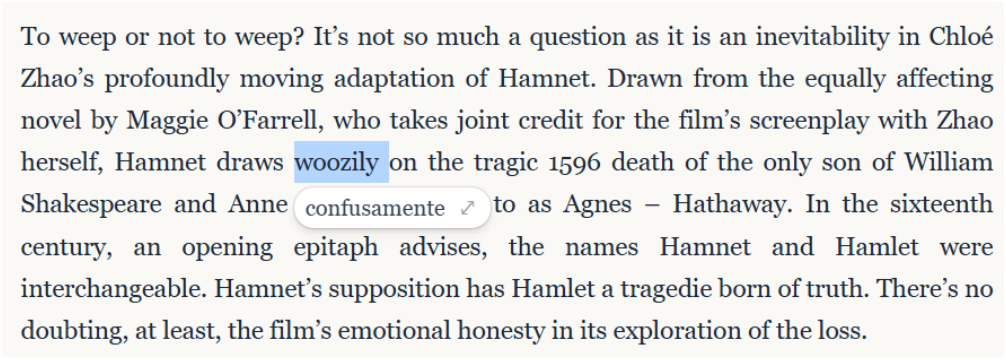



**Figura 6.3:** Schermata di lettura

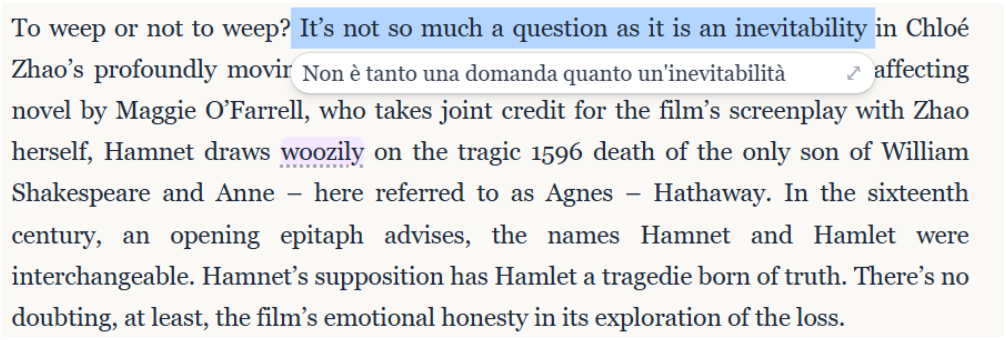
In cima alla pagina si trova un'intestazione fissa, con un pulsante a sinistra per tornare alla libreria e un'icona a destra che dà accesso alle impostazioni. Al momento, le impostazioni permettono solo di scegliere un colore per le note; la selezione della lingua dello strumento, presente in una versione precedente del prototipo e poi rimossa (§6.1), potrebbe eventualmente essere reintrodotta in questa stessa sezione in futuro.

### Selezione e traduzione

Per accedere al primo livello di supporto, l'utente seleziona una qualsiasi porzione di testo, che sia una parola (Figura 6.4a), una frase (Figura 6.4b) o un intero paragrafo. In ogni caso, la traduzione generata dal modello appare in un popover ancorato alla porzione di testo interessata, posizionato sotto di essa o, se lo spazio non è sufficiente, sopra, secondo il principio di minimo ingombro già discusso in §5.1. Accanto alla traduzione si trova il pulsante per passare al pannello di approfondimento. Il popover si chiude cliccando in un punto qualsiasi al di fuori di esso.




**(a)** Selezione di una parola.




**(b)** Selezione di una frase.

**Figura 6.4:** Interazioni del primo livello

### Pannello inline

Aprendo il pannello di approfondimento a partire dal popover di primo livello, questo si inserisce nel testo subito sotto l'ultima riga interessata dalla selezione, spostando verso il basso il contenuto successivo(Figura 6.5). In alto a sinistra, accanto alla dicitura "Esplora come:", compare la categoria assegnata dal modello, presentata come un controllo a tendina cliccabile che rende evidente la possibilità di correggerla (G9, correzione efficiente). Sotto la classificazione sono mostrati i chip contenenti i prompt di approfondimento predefiniti per quella categoria.

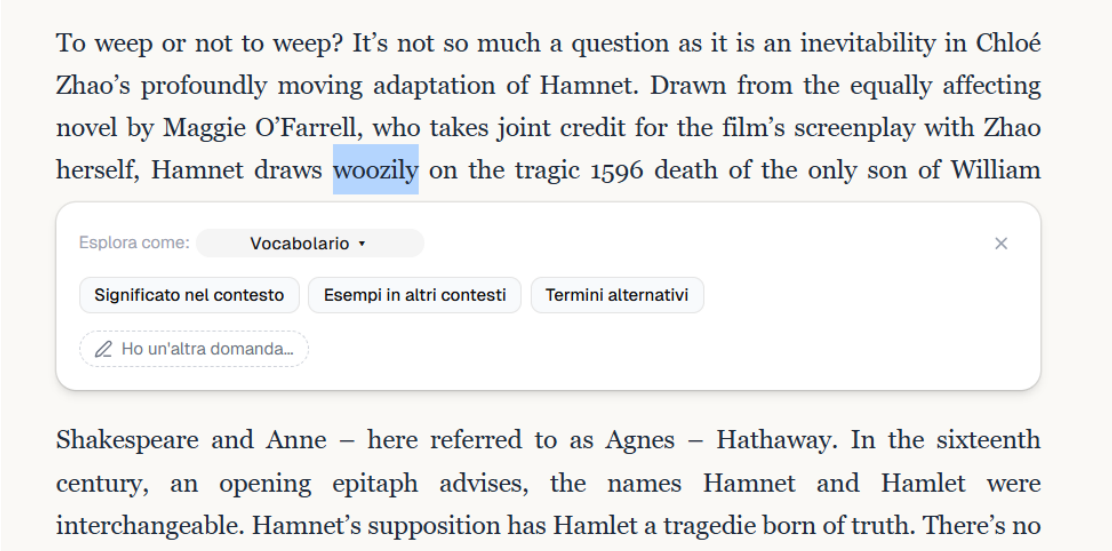



**Figura 6.5:** Pannello inline appena aperto

In fondo a sinistra si trova il pulsante "Ho un'altra domanda", che apre il campo di input per formulare liberamente una domanda (Figura 6.6).

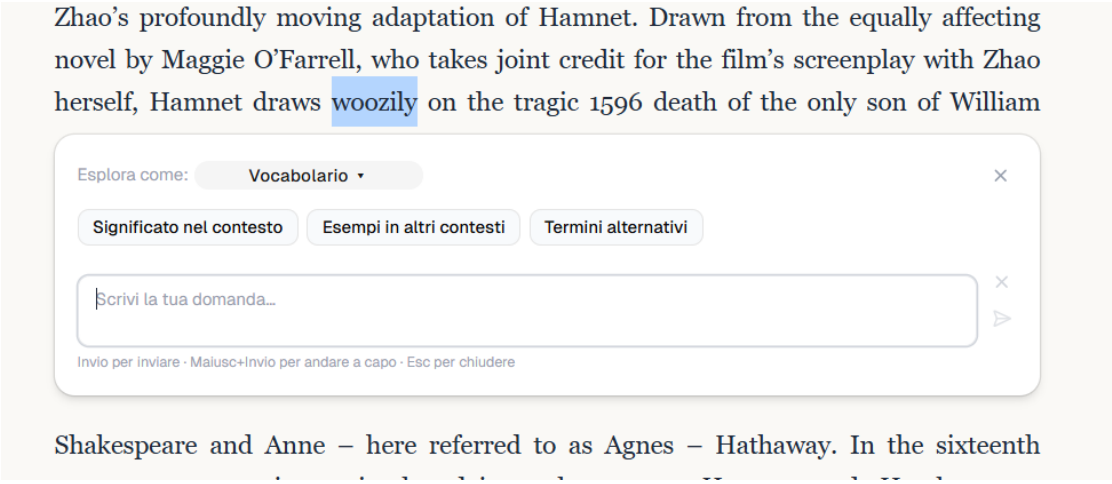



**Figura 6.6:** Apertura input per domanda libera

In alto a destra invece, un pulsante di chiusura permette di chiudere il pannello, così come cliccare in un punto qualsiasi al di fuori di esso.

Cliccando su uno dei chip, viene inviato il prompt di approfondimento corrispondente, e la risposta del modello occupa la maggior parte del pannello (Figura 6.7). Il design iniziale prevedeva che i chip proposti in base alla categoria assegnata

dall'IA scomparissero una volta selezionatone uno, lasciando disponibili solo le domande di approfondimento generate in risposta alla richiesta. Si è deciso invece di mantenere i chip sempre visibili in testa al pannello: trattandosi degli aspetti che l'utente ha più probabilità di voler approfondire, esplorarne uno non deve precludere la possibilità di esplorare anche gli altri.

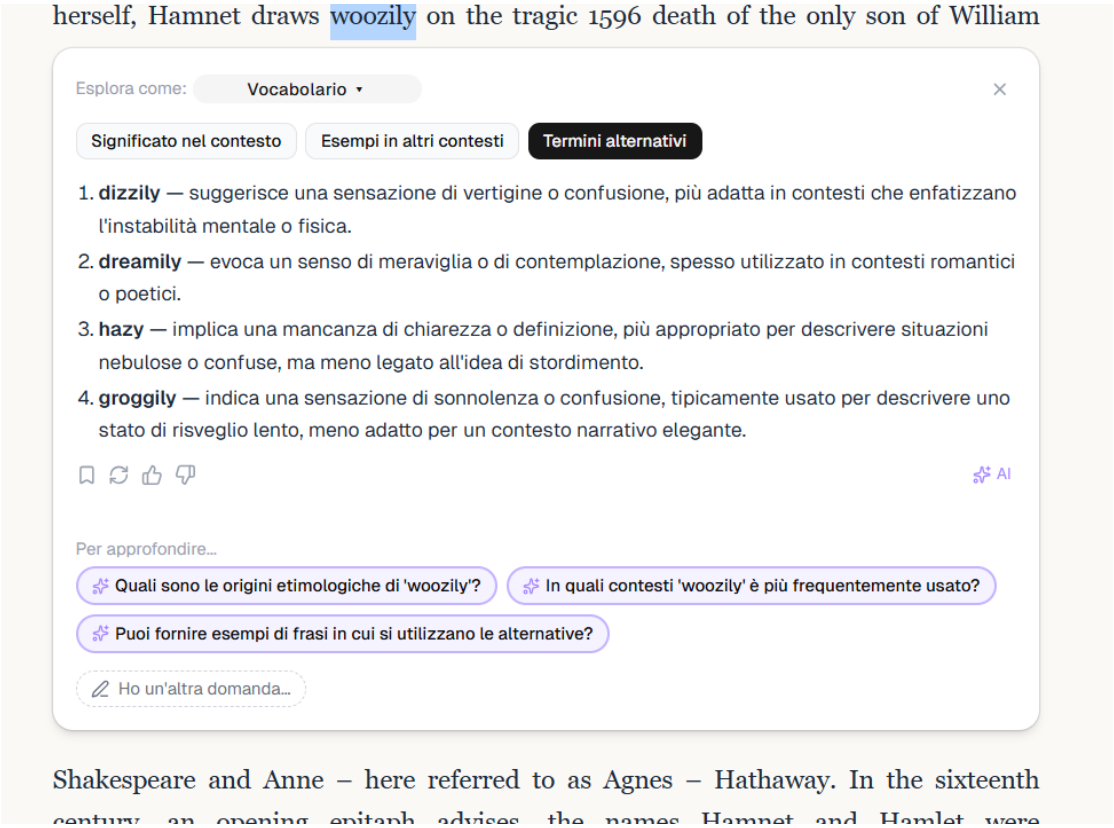



**Figura 6.7:** Generazione risposte in seguito a click su prompt di default

Sotto la dicitura "Per approfondire...", su sfondo colorato, compaiono alcune domande di follow-up che l'utente potrebbe voler porre al modello.

Le domande di follow-up suggerite, così come quelle scritte liberamente dall'utente, una volta inviate compaiono insieme alle relative risposte sotto ai chip fissi, con uno stile visivo leggermente diverso che aiuta a distinguerle da questi ultimi (Figura 6.8) . Quando l'utente ha posto almeno due domande sulla stessa selezione, le risposte ricevute diventano navigabili tramite due frecce sul lato destro del pannello, oltre alla normale barra di scorrimento. Le domande e le risposte generate dal modello sono contrassegnate da un simbolo dedicato, per ricordare all'utente la natura generata del contenuto (G2, comunicazione dell'affidabilità del sistema).




**Figura 6.8:** Seconda interazione tramite domanda di followup

Sotto ciascuna risposta sono disponibili quattro bottoni: un segnalibro, per salvare la risposta nella nota che verrà associata alla porzione di testo selezionata alla chiusura del pannello; un pulsante di rigenerazione, per richiedere una risposta diversa qualora quella ricevuta non fosse soddisfacente (G9, correzione efficiente); due pulsanti a forma di pollice, per indicare apprezzamento o meno della risposta (G15, feedback granulare). Se l'utente esprime un giudizio negativo, può opzionalmente specificare cosa non lo ha soddisfatto in un campo di testo dedicato (Figura 6.9), il cui contenuto viene preso in considerazione nell'eventuale rigenerazione richiesta. Un messaggio conferma che il feedback è stato registrato e sarà utilizzato nelle rigenerazioni successive (G16, comunicazione delle conseguenze delle azioni dell'utente).

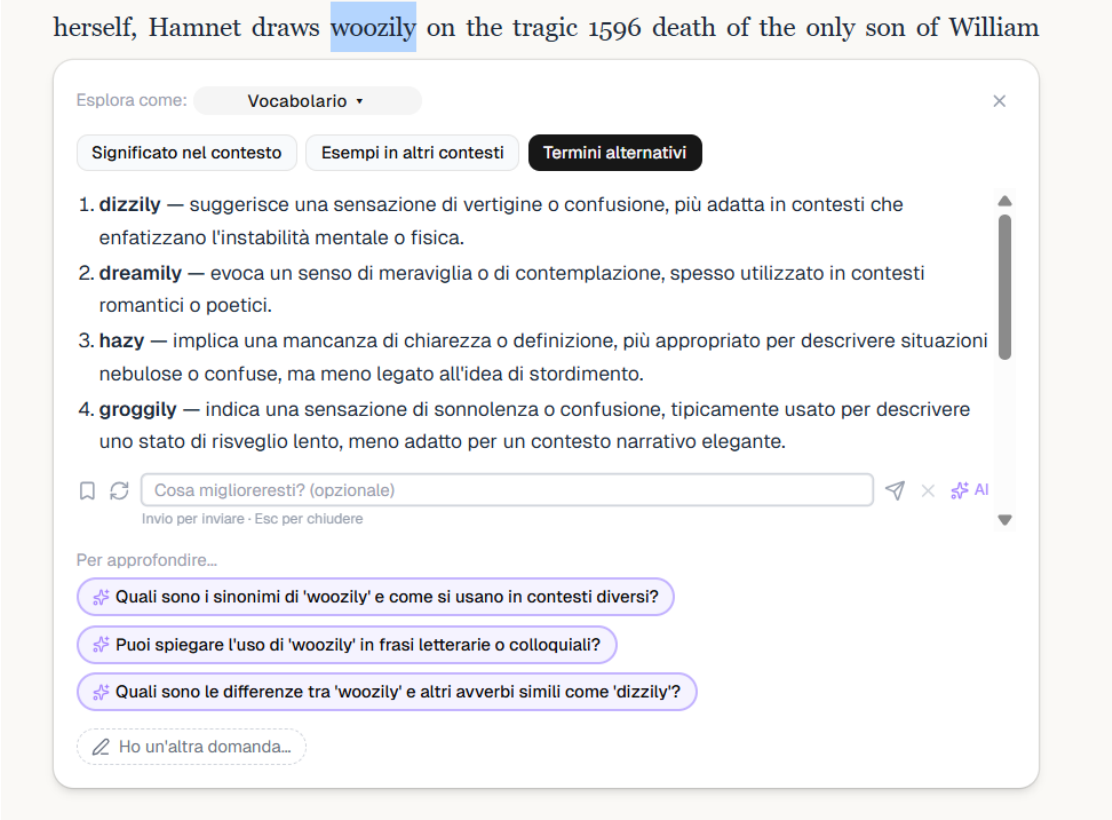



**Figura 6.9:** Possibilità di lasciare un feedback

### Modifica della classificazione

Cliccando sulla categoria assegnata dal modello, si apre un menu a tendina in cui tutte le opzioni disponibili sono mostrate in ordine in base al punteggio di confidenza assegnato internamente dal modello (Figura 6.10). L'utente ha la libertà di selezionare quella che ritiene più adatta, e i chip dei prompt suggeriti cambiano di conseguenza.

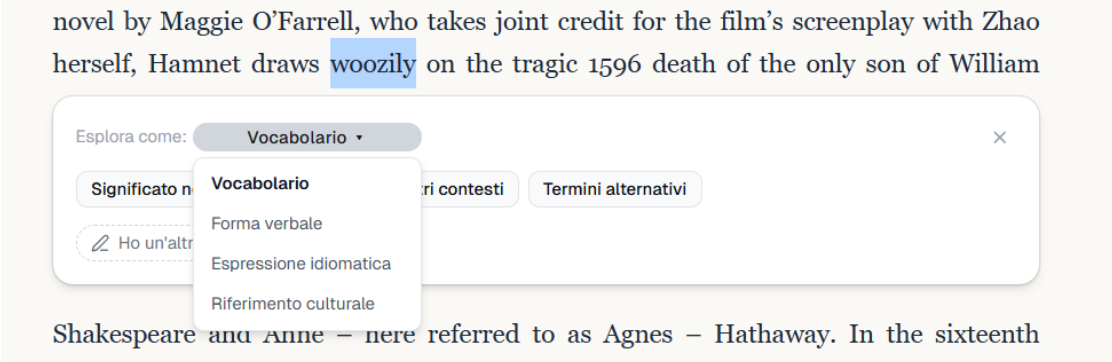



**Figura 6.10:** Menu a tendina per correggere la classificazione

Se il punteggio di confidenza internamente associato alla categoria selezionata dall'utente è inferiore a una certa soglia, un'icona di allerta richiama l'attenzione dell'utente; passandoci sopra con il mouse, un avviso spiega che la categoria selezionata potrebbe non essere pertinente (Figura 6.11). Selezionando comunque uno dei prompt proposti, il modello genera una risposta sulla base della categoria scelta: talvolta il risultato resta adeguato al testo nonostante l'incongruenza segnalata, altre volte il modello stesso evidenzia nella risposta la possibile incompatibilità con il contenuto della selezione.

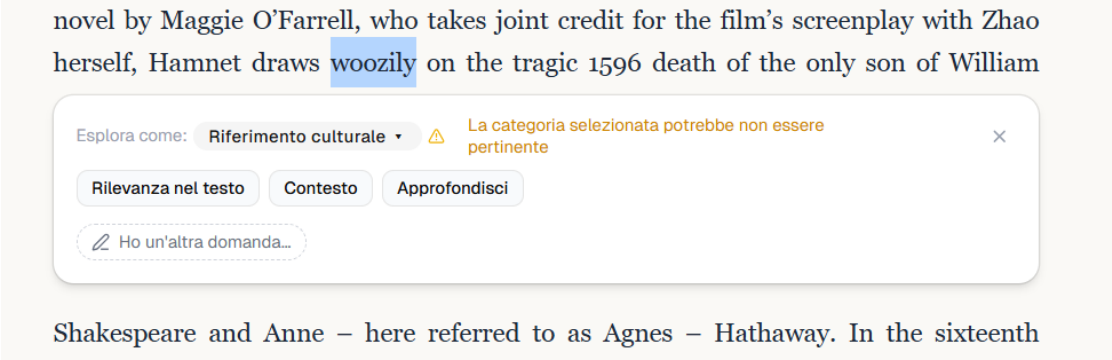



**Figura 6.11:** Avviso di categoria potenzialmente sbagliata

### Note a margine

Una volta conclusa un'interazione con una porzione di testo, che si sia limitata alla traduzione immediata o che sia arrivata all'uso del pannello inline, viene generata una nota associata ad essa, segnalata da una leggera evidenziatura del testo interessato (Figura 6.12).

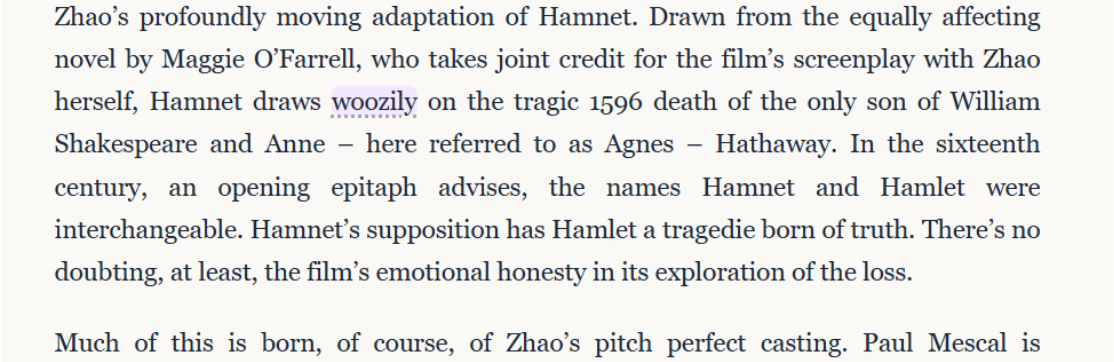



**Figura 6.12:** Il testo con cui l'utente interagisce rimane evidenziato

All'utente è sufficiente cliccare sulla porzione evidenziata per accedere alla nota, che si apre sul lato destro del testo, ancorata alla riga in cui è stata creata (Figura 6.13). Il design definito in fase di progettazione (§5.1) prevedeva che le note comparissero ancorate inline al testo, in linea con il principio di approccio inline discusso in §5.1.1. Una volta implementate in questa forma, è emerso che la compresenza di note e pannello di approfondimento, entrambi inline, rendeva difficile per l'utente distinguerli a colpo d'occhio, generando confusione nel riconoscere quale dei due elementi si stesse per aprire. Si è deciso quindi di mantenere le note ancorate alla porzione di testo a cui si riferiscono, ma di posizionarle sempre lateralmente rispetto ad essa. Questa scelta non è in conflitto con il principio di approccio inline: la nota risponde a un bisogno che si manifesta tipicamente in un momento successivo alla lettura, e la sua apertura, per quanto ora laterale, resta comunque un'interazione rapida e reversibile con un solo click.

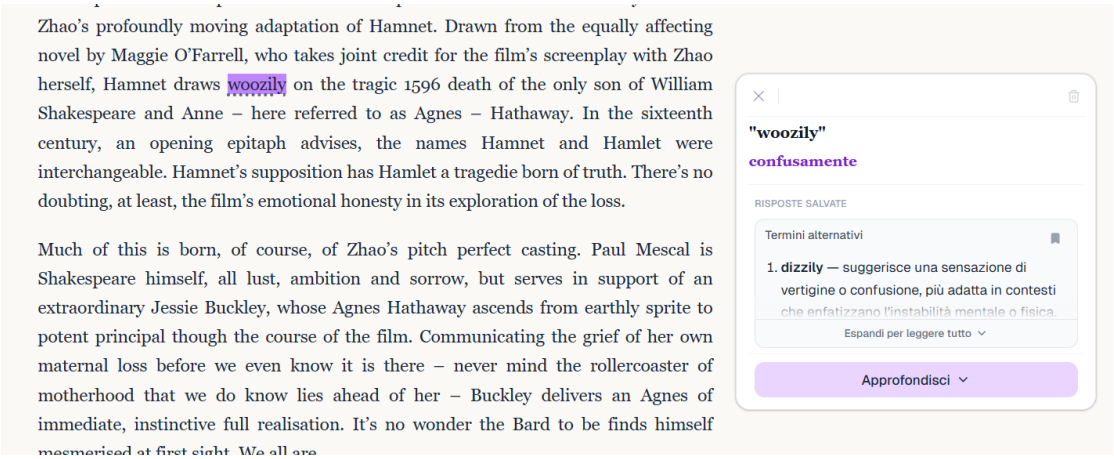



**Figura 6.13:** La nota si apre a lato del testo

All'interno della nota è riportata la porzione di testo interessata, con la traduzione corrispondente immediatamente sotto. Seguono eventuali risposte salvate durante l'approfondimento avvenuto nel pannello, visionabili in forma estesa cliccando su un apposito pulsante (Figura 6.14).

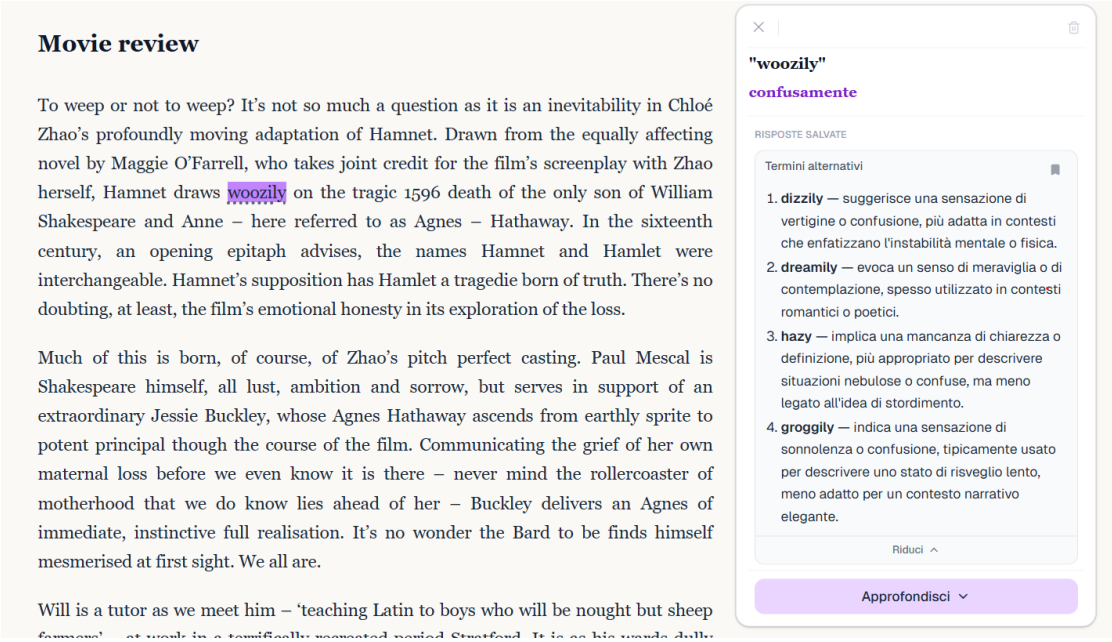



**Figura 6.14:** Apertura di risposte ricevute in passato nella nota

In fondo alla nota, un pulsante denominato "Approfondisci" permette all'utente di aprire il pannello inline in corrispondenza della porzione di testo interessata, che conterrà lo storico delle interazioni già avvenute, se presente.

### 6.3.4 Componenti principali

L'organizzazione dei componenti segue la distinzione tra livelli introdotta in §6.3.1: i componenti di pagina si occupano del caricamento dei dati necessari e delegano il resto; i componenti orchestratori compongono gli hook e decidono cosa mostrare; i restanti componenti sono presentazionali, e ricevono dati e funzioni tramite proprietà senza gestire stato o effettuare chiamate proprie.

**Reader.** È il componente orchestratore principale della vista di lettura, e il più corposo dell'applicazione. Compone gli hook responsabili della selezione testuale, delle annotazioni e dello stato generale della sessione di lettura, e gestisce l'intestazione fissa della pagina. Il punto in cui inserire il pannello di approfondimento

all'interno di un paragrafo è calcolato da uno degli hook di gestione dello stato di lettura; Reader riceve questa posizione già determinata e si occupa di suddividere di conseguenza il rendering del paragrafo interessato. Contiene inoltre direttamente, senza delegarla a un hook, la logica di evidenziazione delle porzioni di testo annotate, incluso il caso in cui più annotazioni finiscano per sovrapporsi tra loro.

**SelectionPopover.** Realizza il popover del primo livello di supporto (§6.3.3): è ancorato alla porzione di testo selezionata, e mostra un'indicazione di caricamento mentre la traduzione viene generata, seguita dal risultato una volta ricevuto, insieme al pulsante per accedere al livello di approfondimento successivo.

**LevelTwoPanel.** Realizza il pannello di approfondimento (§6.3.3). È composto da un'intestazione, che mostra e permette di correggere la categoria assegnata dal sistema (§6.3.3), e da un corpo, che raccoglie i prompt suggeriti, l'input per la domanda libera, i prompt di approfondimento generati dal modello e lo storico delle interazioni già ricevute su quella selezione. Ciascuno di questi blocchi è a sua volta scomposto in componenti presentazionali più piccoli, dedicati a un singolo elemento di interfaccia

**AnnotationNote.** Realizza la nota persistente (§6.3.3): mostra la risposta immediata salvata insieme alle interazioni contrassegnate come salvate dall'utente, e permette di riaprire il pannello di approfondimento sulla stessa selezione. Espone due azioni distinte di cancellazione: l'eliminazione dell'intera nota, che richiede una conferma esplicita da parte dell'utente prima di procedere, e la rimozione di una singola interazione dall'elenco di quelle salvate, azione più leggera che non richiede conferma.

## 6.4 Backend

Questa sezione descrive come le componenti lato server del sistema sono state implementate, a partire dalle tecnologie impiegate fino all'organizzazione interna del codice e alle singole responsabilità funzionali, tra cui la generazione dei contenuti, la gestione della conversazione e la persistenza dei dati.

### 6.4.1 Architettura e organizzazione

Il backend è organizzato secondo un'architettura a livelli, che separa la gestione delle richieste HTTP dalla logica di business e dall'accesso ai dati. Questa separazione permette di modificare o testare ciascun livello in modo relativamente indipendente dagli altri.

Un modulo dedicato, `router.ts`, mappa ogni endpoint REST alla funzione responsabile di gestirlo, senza contenere alcuna logica applicativa propria. Da lì, ogni richiesta segue un flusso lineare attraverso tre livelli principali: *controller*, *service* e *repository*. I *controller* costituiscono l'unico punto di contatto diretto con le richieste e risposte HTTP: validano i dati ricevuti tramite schemi definiti con Zod, e delegano la logica effettiva al *service* corrispondente. Gli stessi schemi Zod sono riutilizzati anche per vincolare la struttura delle risposte generate dal modello linguistico, garantendo che rispettino un formato prevedibile. I *service* contengono la logica di business vera e propria: costruiscono i prompt da inviare al modello, gestiscono lo storico della conversazione e orchestrano le chiamate necessarie, senza mai accedere direttamente al database. Questo compito è riservato ai *repository*, unico livello a interfacciarsi con Prisma e, tramite esso, con il database PostgreSQL descritto in §6.2.

L'integrazione con il modello linguistico è isolata in un client dedicato, un unico punto di accesso all'SDK di OpenAI usato da tutti i *service*: nessuna altra parte del codice comunica direttamente con il provider. Questo isolamento rende più agevole, in prospettiva, un'eventuale sostituzione del modello o del provider utilizzato.

### 6.4.2 Generazione dei contenuti

Le tre funzionalità generative del sistema sono esposte tramite altrettanti endpoint, e condividono lo stesso meccanismo di base: ogni chiamata al modello linguistico utilizza l'output strutturato offerto dall'API di OpenAI, che vincola la risposta a rispettare uno schema Zod definito a priori. Il service non deve quindi effettuare alcun parsing manuale della risposta del modello, riducendo il rischio di errori di interpretazione.

**Risposta immediata (POST /api/minimal).** Riceve la porzione di testo selezionata insieme al contesto circostante, senza alcuna classificazione preventiva, e restituisce una traduzione calibrata sul significato specifico che il termine assume in quel contesto, non su una definizione generica. Se la selezione supera una soglia di lunghezza, il sistema utilizza una variante del prompt pensata per gestire selezioni multi-frase o multi-paragrafo; sotto soglia, viene utilizzata un'istruzione più diretta.

**Classificazione (POST /api/classify).** Al di sopra della stessa soglia di lunghezza appena descritta, la classificazione viene assegnata direttamente alla categoria "passaggio confuso" senza interpellare il modello, evitando una chiamata quando l'esito è ragionevolmente prevedibile. Al di sotto della soglia, il modello classifica la selezione tra le categorie rimanenti, restituendo anche un punteggio di confidenza per ciascuna. I prompt suggeriti mostrati all'utente non sono generati dal modello, ma recuperati da una tabella statica definita per ciascuna categoria:

la scelta di quali approfondimenti proporre resta quindi un contenuto controllato, non soggetto a variazione tra una richiesta e l'altra.

**Approfondimento (POST /api/respond).** È l'endpoint con la logica più articolata, poiché la costruzione del messaggio inviato al modello varia in base alla combinazione tra categoria e prompt scelto dall'utente. Per il riassunto, ad esempio, non viene inviato alcun contesto circostante, ma solo il testo selezionato, per evitare che il riassunto derivi dal contesto anziché dalla porzione di testo effettivamente indicata dall'utente. In ogni richiesta viene inoltre inclusa un'istruzione che invita il modello ad adattarsi al contenuto reale del testo qualora la categoria assegnata risultasse in realtà poco pertinente, introducendo un margine di correzione rispetto a un possibile errore nel passo di classificazione precedente. Oltre al testo della risposta, il modello restituisce alcune domande di approfondimento opzionali, proposte come prompt suggeriti aggiuntivi nel pannello.

Per garantire continuità nelle interazioni successive sulla stessa selezione, viene recuperato e fornito al modello lo storico della conversazione associata alla selezione; la gestione di questo storico è descritta in dettaglio in §6.4.3. Il sistema supporta inoltre la rigenerazione di una risposta già ricevuta, sulla base della risposta precedente ed eventuale feedback fornito dall'utente.

La scrittura dei singoli prompt ha seguito alcuni principi trasversali. Per esempio, ciascun prompt assegna esplicitamente al modello un ruolo di assistente alla lettura in L2, per ancorarne il tono e il registro delle risposte. Per i formati di output più strutturati, come le liste di esempi o di termini alternativi, è stato inoltre impiegato il few-shot prompting, fornendo al modello uno o più esempi del formato atteso per ridurre la variabilità nella struttura della risposta.

### 6.4.3 Gestione delle conversazioni

Una volta che l'utente ha iniziato a interrogare una porzione di testo, lo scambio di domande e risposte ad essa relativo viene mantenuto nel tempo, in modo da dare al modello una visione complessiva della conversazione, evitando che ogni nuova interazione venga trattata ex novo e che vengano ripetute informazioni già fornite. Questo è ottenuto assegnando a ciascuna conversazione un identificativo generato in modo casuale al momento del primo scambio, associato a ciascun messaggio successivo relativo alla stessa porzione di testo.

Prima di generare una nuova risposta, il sistema verifica se esiste già uno storico per quella conversazione e, in caso affermativo, lo fornisce al modello come sequenza di messaggi con ruolo e contenuto, ricostruendo manualmente la cronologia a ogni richiesta: l'API del modello linguistico non mantiene infatti alcuno stato tra le chiamate, ed è quindi il backend a recuperare i messaggi precedenti dal database e a includerli nella richiesta corrente insieme al nuovo messaggio

dell'utente. L'assenza di uno storico pregresso distingue una prima domanda da una domanda di approfondimento. Nel primo caso il messaggio inviato al modello include il testo selezionato, il contesto circostante e la categoria assegnata, mentre, nei turni successivi, poiché queste informazioni sono già presenti nello storico, viene inviata solo la nuova domanda. Al termine di ogni scambio andato a buon fine, i nuovi messaggi vengono aggiunti alla conversazione.

Anche una rigenerazione genera un nuovo scambio che viene aggiunto alla stessa conversazione, così che il modello abbia visibilità, nei turni successivi, anche dei tentativi di risposta rigenerati, non solo dell'ultimo accettato. Quando l'utente fornisce un feedback sulla risposta da rigenerare, questo viene incluso esplicitamente nella richiesta inviata al modello insieme alla risposta precedente, con l'istruzione di produrne una diversa e migliorata tenendone conto. A livello del pannello visto dall'utente, tuttavia, la rigenerazione aggiorna l'interazione già esistente invece di crearne una nuova e l'utente continua a vedere una sola risposta per quella richiesta, anche se la cronologia interna della conversazione cresce a ogni tentativo.

### 6.4.4 Persistenza dei dati

La persistenza dei dati risponde direttamente al bisogno di ritrovare nel tempo le interazioni avute con il testo, discusso in §5.1.5. Il modello dati, illustrato in Figura 6.14, riflette questa priorità: separa il contenuto letto, le annotazioni create durante la lettura, le singole interazioni che le compongono, e lo storico conversazionale che le sostiene, in modo che ciascun elemento possa essere gestito e recuperato in modo indipendente.

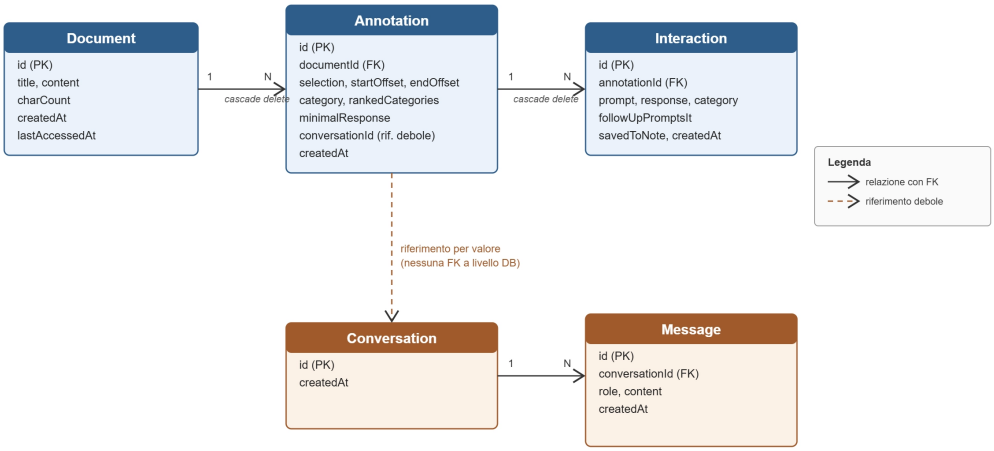



**Figura 6.15:** Diagramma entità-relazione del modello dati.

**Documenti.** Ogni testo caricato dall'utente è salvato come una singola riga (POST /api/documents), contenente il testo per intero insieme a titolo e conteggio dei caratteri. Non è previsto il caricamento di file, in quanto il contenuto è trattato come testo semplice, salvato direttamente nel database. Ogni apertura di un documento (GET /api/documents/:id) aggiorna la data di ultimo accesso, così da poter ordinare la lista dei documenti recenti (GET /api/documents).

**Annotazioni e interazioni.** Ogni annotazione corrisponde a una porzione di testo selezionata dall'utente, identificata dalla propria posizione nel documento, e viene creata (POST /api/annotations) già al momento della prima risposta immediata, prima ancora che l'utente scelga di approfondire. Ogni domanda posta nel pannello di approfondimento, con la relativa risposta ricevuta, costituisce un'interazione, collegata all'annotazione a cui si riferisce; un'annotazione può quindi raccogliere più interazioni via via che l'utente approfondisce la stessa porzione di testo. Cancellare un'annotazione (DELETE /api/annotations/:id) elimina automaticamente tutte le interazioni ad essa collegate. Ciascuna interazione può inoltre essere contrassegnata come salvata o meno (PATCH /api/interactions/:id/saved): questo determina se compare nella nota persistente, senza comportare la creazione o l'eliminazione di alcun dato.

**Conversazioni.** Lo storico multi-turno descritto in §6.4.3 è mantenuto in una struttura dati indipendente da quella delle annotazioni: un'annotazione vi fa riferimento tramite un identificativo condiviso, senza che esista un vincolo relazionale esplicito a livello di database tra le due strutture. Questo è coerente con il comportamento previsto dal design: cancellando un'annotazione, l'utente si aspetta che anche lo storico delle interazioni ad essa associate non sia più raggiungibile, e infatti una nuova selezione della stessa porzione di testo genera una nuova conversazione, indipendente dalla precedente. A livello di database, tuttavia, le righe corrispondenti allo storico eliminato non vengono effettivamente rimosse, ma restano come dati orfani, non più raggiungibili da alcun riferimento. Questa mancata pulizia dei dati è un limite tecnico del prototipo, discusso in §8.1.

Infine, non è prevista alcuna forma di autenticazione o distinzione tra utenti: il prototipo è stato sviluppato come applicazione single-user. Gli adattamenti resi necessari dalla conduzione dello studio a più partecipanti sono descritti in §7.1.3.

# Capitolo 7

# Valutazione

Questo capitolo descrive lo studio di valutazione condotto per quantificare il contributo di ReadLens alla comprensione e alla ritenzione di lessico inglese durante la lettura in L2. Vengono presentati di seguito il protocollo adottato, i risultati ottenuti e la loro discussione alla luce dei dati quantitativi e qualitativi raccolti.

## 7.1 Metodo

Il disegno dello studio si colloca deliberatamente tra due tradizioni. La prima, quella degli studi controllati sul glossing, isola tipicamente l'effetto dello strumento sull'acquisizione di parole target predefinite tramite test somministrati prima e dopo la lettura, ma non osserva l'interazione reale con lo strumento.

La seconda, quella dei lavori HCI su strumenti di apprendimento linguistico in contesto, privilegia deployment studies naturalistici e analisi delle interazioni, ma raramente cerca di misurare rigorosamente il contributo all'apprendimento. Questo protocollo mira a combinare le due tradizioni, unendo misure standard del campo SLA a un logging completo delle interazioni.

Lo studio adotta un disegno within-subjects, con due condizioni sperimentali, ReadLens e uno strumento di controllo, Baseline, e due passaggi di lettura.

### 7.1.1 Domande di ricerca

A inizio studio sono state stabilite cinque domande di ricerca che hanno guidato la definizione del protocollo e la successiva analisi e interpretazione dei dati. L'ultima è dichiarata sin dall'inizio come esplorativa.

1. RQ1. Lo strumento migliora la comprensione di testi inglesi rispetto a un baseline LLM-powered standard?

2. RQ2. Lo strumento favorisce l'acquisizione incidentale di vocabolario, immediata e ritardata, rispetto al baseline?

3. RQ3. Come si distribuisce l'engagement degli utenti sui due livelli di progressive disclosure, e per quali categorie di difficoltà?

4. RQ4. Lo strumento è percepito come meno interrompente e meno faticoso del baseline?

5. RQ5 (esplorativa). La profondità di esplorazione raggiunta su un item correla con la sua retention ritardata?

### 7.1.2 Partecipanti

Lo studio ha coinvolto 20 partecipanti, di cui due impiegati come pilot testers della procedura di valutazione e per questo non considerati nelle analisi dei risultati. Le adesioni sono state raccolte tramite sondaggio su Google Form, diffuso attraverso canali social e gruppi studenteschi del Politecnico.

Il sondaggio ha permesso di filtrare le risposte, considerando solo chi rispettasse i criteri di inclusione prestabiliti: lingua italiana come L1, inglese di livello B2-C1 e un'abitudine di lettura frequente in inglese da computer, autodichiarata.

Non è stata permessa la partecipazione a persone che avessero già preso parte alle interviste dello studio formativo, per evitare che fossero già a conoscenza dello scopo dello studio e che ciò influenzasse i dati. Inoltre sono stati esclusi individui bilingui o con formazione specialistica in linguistica inglese, traduzione o interpretariato.

La partecipazione è stata su base volontaria e senza compenso.

### 7.1.3 Materiali

**Passaggi di lettura:** sono stati scelti due passaggi autentici di lunghezza simile, tra le 800 e le 900 parole, tratti da due articoli di viaggio della BBC[1].

La selezione ha cercato di escludere testi che contenessero un numero eccessivo di nomi propri o richiedessero conoscenze pregresse per essere compresi. Per verificarne la comparabilità sono stati calcolati gli indici di leggibilità Flesh-Kincaid

[1]Passaggio A: *Photographing one of the most remote places on Earth*, BBC Travel, disponibile su `https://www.bbc.com/travel/article/20240816-ittoqqortoormiit-the-most-remote-community-in-the-western-hemisphere` (consultato il 5 Luglio 2026). Passaggio B: *'It's like being in Greece': The US neighbourhood where cars are banned*, BBC Travel, disponibile su `https://www.bbc.com/travel/article/20250815-the-us-neighbourhood-where-cars-are-banned` (consultato il 5 Luglio 2026).

[28](tramite Microsoft Word), e il profilo di frequenza lessicale con VocabProfilers Compleat di Lextutor, che classifica il vocabolario secondo le liste di famiglie lessicali BNC/COCA definite da Nation [29]. Pur non essendo identici, i due testi sono risultati la coppia con le metriche più simili tra quelle riscontrate nella ricerca condotta, infatti fino alla soglia K-5 le distribuzioni sono molto simili, con differenze isolate solo ai livelli di frequenza più alti.

Per selezionare le parole target la cui ritenzione sarebbe stata osservata durante lo studio, i due testi sono stati fatti leggere a quattro studenti universitari, compatibili con i criteri di inclusione dello studio e reclutati tramite passaparola, di cui due di livello B2 e due C1, a cui è stato chiesto di sottolineare le parole o le espressioni con cui riscontravano una difficoltà di comprensione. Per ciascun testo sono state selezionate solo le parole segnalate da almento due dei quattro lettori, ottenendo due liste di dodici parole o espressioni target ciascuna.

I due insiemi sono risultati ragionevolmente bilanciati per frequenza BNC/COCA, senza grosse disparità tra loro.

**ReadLens:** il prototipo usato come una delle due condizioni sperimentali è quello risultante dalla fase di implementazione, descritto nel Capitolo 6. Per la conduzione dello studio, è stato esteso con alcuni adattamenti specifici. È stato aggiunto un meccanismo di autenticazione basato su token per cui a ciascun partecipante viene associato un token univoco, scambiato con un cookie di sessione all'accesso tramite un link personalizzato, che protegge le rotte in scrittura del sistema, accompagnato da controlli di ownership che impediscono a un partecipante di accedere ai dati di un altro. Il percorso verso la libreria dei documenti, che nel prototipo di base permette il caricamento libero di un testo, è stato inoltre sostituito da un accesso diretto ai due passaggi predefiniti per lo studio

**Baseline:** Il Baseline è stato implementato in modo tale da rendere disponibili durante la lettura una funzione di dizionario, attivabile con un doppio click su una parola per mostrarne la definizione in un tooltip, e una chat laterale collegata a un modello linguistico, istruita con un prompt da assistente di lettura.L'inclusione di un accesso a un LLM nel Baseline è stata una scelta deliberata poiché senza di essa, lo studio avrebbe finito per misurare l'effetto dell'accesso a un modello linguistico in sé, anziché il contributo specifico del design e delle interazioni proprie di ReadLens.

### 7.1.4 Strumenti di misura

Per rispondere alle domande di ricerca sono stati impiegati diversi strumenti, che hanno permesso la raccolta di dati sia quantitativi che qualitativi.

**Vocabulary Levels Test (VLT).** Il test è stato somministrato prima delle sessioni principali dello studio, con due funzioni: verificare che ciascun partecipante possedesse una base di competenza linguistica sufficiente a non trovarsi in eccessiva difficoltà nella lettura dei testi, e ottenere una misura della sua competenza da utilizzare come covariata nell'analisi. È stata usata la versione aggiornata di Webb, Sasao e Ballance (2017) [30], di cui sono state selezionate le bande 3K, 4K e 5K. Come criterio di esclusione sono stati accettati solo i partecipanti che avessero risposto correttamente ad almeno 29 delle 30 domande della banda 3K, punteggio indicato dagli stessi autori del test come soglia di padronanza per i primi tre livelli. La banda 3K è stata scelta perché indicativamente associata al livello CEFR B2.

**Vocabulary Knowledge Scale (VKS).** Il VKS nasce come test per la valutazione del riconoscimento isolato del vocabolario, permettendo di valutarne la conoscenza secondo cinque livelli, dal semplice riconoscimento alla dimostrazione di saper utilizzare il termine. La versione utilizzata è quella presentata da Paribakht e Wesche (1996) [31], riformulata in forma breve: il test non prevede l'ultimo livello, che richiede la produzione di una frase contenente il termine, per contenere i tempi di somministrazione. Questa scelta è ispirata al lavoro di Khezrlou et al. [6], dove vengono impiegati i soli livelli ricettivi.

**Word Recognition Test (WRT).** Il WRT valuta la capacità di riconoscere il significato di un termine all'interno del contesto in cui è inserito. Viene fornita una frase contenente il termine, insieme a quattro opzioni di significato tra cui scegliere; i distrattori sono stati costruiti nella stessa area semantica delle parole target. Il formato del test è stato costruito ad hoc per questo studio, ispirandosi all'approccio adottato da Khezrlou et al [6].

**Multiple Choice Question (MCQ).** Un test a scelta multipla costruito ad hoc per ciascun passaggio, con sei domande: tre letterali, che chiedono un'informazione esplicitamente presente nel testo, e tre inferenziali, che richiedono di dedurre qualcosa non detto direttamente, come il significato implicito di una frase o la motivazione di un'affermazione dell'autore.

**Nasa-TLX.** Per rilevare il carico percepito associato all'attività di lettura assistita, è stata usata la versione Raw TLX del test definito da Hart e Staveland (1988) [32], costituita da sei scale che valutano rispettivamente richiesta mentale, fisica e temporale, prestazione, sforzo e frustrazione percepiti dal partecipante nello svolgimento del compito.

**System Usability Scale (SUS).** Misura l'usabilità percepita del sistema appena utilizzato. La versione impiegata è la versione italiana del System Usability Scale riportata da Borsci et al (2009)[33], costituita dai medesimi dieci item del SUS originariamente proposto da Brooke (1996)[34], tradotti e in alcuni casi adattati nella formulazione italiana. Il test chiede al partecipante di indicare il proprio grado di accordo con una serie di affermazioni su una scala Likert a 5 punti, da "fortemente in disaccordo" a "fortemente d'accordo".

**Intervista semi-strutturata.** Al termine della prima sessione, i partecipanti hanno sostenuto un'intervista semi-strutturata, nella quale è stato chiesto loro di raccontare la propria esperienza ed esprimere osservazioni sui due strumenti. Le domande utilizzate come traccia indicativa sono le seguenti:

1. Raccontami com'è andata. Come sono state le due letture? Le hai trovate diverse o abbastanza simili?
2. In alcuni momenti hai approfondito di più, in altri ti sei fermata/o prima o non hai usato nulla. Cosa guidava quella scelta?
3. Gli strumenti hanno cambiato il tuo modo di leggere? Ti hanno mai fatto perdere il filo del discorso o distratto?
4. C'è qualcosa che ti sarebbe servito e che nessuno dei due strumenti ti dava?
5. Le spiegazioni erano in italiano. Come ti è sembrato? C'è stato un momento in cui avresti preferito diversamente?
6. Dovendo sceglierne uno per le tue letture in inglese di tutti i giorni, quale sceglieresti? Perché?

**Logging delle interazioni.** Durante entrambe le condizioni, il sistema ha registrato automaticamente il progresso di lettura di ciascun partecipante insieme al tempo impiegato.

In condizione ReadLens, il sistema ha inoltre registrato ogni interazione del partecipante con lo strumento: selezioni di testo, livello di approfondimento raggiunto per ciascuna, categoria assegnata dalla classificazione automatica, domande poste e risposte ricevute, apertura e chiusura delle note. Questi dati hanno permesso di ricostruire, per ciascun partecipante, il comportamento di lettura effettivo, alla base delle analisi su engagement (RQ3) e sulla profondità di esplorazione raggiunta per singolo item (RQ5).

In condizione Baseline sono stati registrati l'utilizzo del dizionario integrato e le domande poste tramite la chat, ma i due strumenti sono tracciati come eventi indipendenti tra loro.

### 7.1.5 Procedura

La procedura prevedeva tre sessioni distinte per partecipante. Per accedere a ciascuna, il partecipante doveva aprire il link ricevuto alla conferma della propria registrazione allo studio: lo stesso link dava accesso, per tutta la durata dello studio, all'applicazione web che ospitava i vari passaggi, distribuita tramite Heroku per consentirne la fruizione da remoto.

La prima sessione, svolta in remoto e in autonomia, consisteva unicamente nel test VLT, usato come screening per verificare che il livello autodichiarato dal partecipante fosse coerente con il punteggio minimo richiesto per partecipare al resto dello studio.

La seconda sessione, la più impegnativa, si svolgeva via videochiamata su Microsoft Teams e prevedeva la lettura di entrambi i passaggi, uno per ciascuna condizione sperimentale.

La sessione iniziava con un breve benvenuto e la lettura del consenso informato. Seguiva il pre-test VKS, che valutava la conoscenza pregressa dei 24 item target selezionati dai due passaggi, prima che questi fossero letti.

Il partecipante passava poi a una schermata di onboarding, contenente un testo di prova e lo strumento che avrebbe utilizzato nella prima lettura. Questa schermata serviva a esporre le funzionalità dello strumento prima del momento di lettura di interesse per lo studio, per non inquinare la raccolta dei dati sull'interazione spontanea né forzare l'interazione con una possibile parola target; è stato infatti verificato a priori che il testo di prova non contenesse alcuna delle parole appartenenti agli item target dell'analisi sulla ritenzione. Una volta che il partecipante dichiarava di aver compreso il funzionamento dello strumento, procedeva alla prima lettura.

Le istruzioni fornite a inizio lettura erano le seguenti: "Ti chiedo di leggere questo testo con l'obiettivo di capirlo, come faresti normalmente. Hai a disposizione lo strumento di supporto che ti ho appena mostrato: sei libera/o di usarlo se e quando ti serve, oppure di non usarlo affatto. Non c'è un tempo limite rigido; indicativamente la lettura richiede 15 minuti." La formulazione evitava deliberatamente di indicare esplicitamente quando ricorrere allo strumento, poiché alcuni studi mostrano che istruzioni esplicite sull'uso dei gloss tendono ad aumentare l'apprendimento a parità di strumento, rispetto alla lettura senza indicazioni [6]. Per coerenza con l'obiettivo di osservare un apprendimento incidentale, si è optato per questa formulazione più conservativa.

Al termine della lettura, il partecipante valutava il carico percepito dell'attività tramite Nasa-TLX e l'usabilità dello strumento tramite SUS. Seguivano poi l'MCQ, relativo al testo appena letto, un VKS contenente i soli 12 item target dello stesso testo, per valutare l'eventuale guadagno nella ritenzione immediata, e infine il WRT sugli stessi 12 item.

Il processo si ripeteva quindi per la seconda condizione e il secondo testo, a partire dalla schermata di onboarding relativa al secondo strumento.

I partecipanti, all'atto del consenso informato, venivano assegnati a uno di quattro gruppi, la cui appartenenza determinava l'ordine di fruizione dei due strumenti e la loro corrispondenza con i due testi. I gruppi erano organizzati secondo un disegno a quadrato latino, che bilancia eventuali effetti dovuti all'ordine delle condizioni e dei testi, come il testing effect che potrebbe derivare dall'aspettarsi, dopo la prima lettura, un testo e un compito analoghi anche nella seconda, modificando di conseguenza il proprio comportamento. I partecipanti non erano informati a priori della struttura delle tre sessioni, per non rischiare di alterarne i comportamenti naturali, già potenzialmente condizionati dal dover svolgere la sessione in videochiamata con condivisione dello schermo.

La sessione si concludeva con le domande dell'intervista semi-strutturata previste dal protocollo, volte a indagare l'esperienza e le impressioni del partecipante. La conversazione veniva registrata e la registrazione cancellata al termine della trascrizione. La sequenza completa della sessione è sintetizzata in Tabella 7.1

| **Step** | **Attività** |
|---|---|
| 1 | Benvenuto e consenso informato |
| 2 | Pre-test VKS (24 item target) |
| 3 | Onboarding con testo di prova sulla prima condizione |
| 4 | Prima lettura (Readlens o Baseline, secondo il quadrato latino) |
| 5 | Nasa-TLX, SUS, MCQ, VKS (12 item), WRT (12 item) |
| 6 | Onboarding con testo di prova sulla seconda condizione |
| 7 | Seconda lettura (condizione rimanente) |
| 8 | Nasa-TLX, SUS, MCQ, VKS (12 item), WRT (12 item) |
| 9 | Intervista semi-strutturata |

**Tabella 7.1:** Sequenza delle attività nella seconda sessione dello studio.

La terza sessione si sbloccava sette giorni dopo il completamento della seconda, e poteva essere svolta da remoto e in autonomia entro i tre giorni successivi, per contenere il rischio di abbandono dello studio. Significativamente più breve delle precedenti, prevedeva un VKS e un WRT, entrambi comprensivi di tutti e 24 gli item target, con l'obiettivo di misurare la ritenzione ritardata dei termini incontrati durante la lettura. Durante questa sessione, il sistema registrava automaticamente eventuali tempi di compilazione anomali per singolo item e segnali di cambio di schermata, come indicatori indiretti di una possibile consultazione di risorse esterne durante il test.

Ciascuna delle tre sessioni era preceduta da una pagina di consenso informato. Al termine dell'ultima sessione, ai partecipanti veniva spiegato nel dettaglio cosa lo studio avesse misurato e con quale scopo, offrendo la possibilità di ritirarsi e richiedere la cancellazione dei propri dati.

### 7.1.6 Analisi dei dati

Le analisi statistiche sono state pianificate prima dell'inizio della raccolta dati, con eventuali scelte metodologiche successive dichiarate esplicitamente dove rilevante.

Per le misure a singolo punteggio per partecipante e condizione (MCQ, Nasa-TLX, SUS, tasso di approfondimento), il confronto tra le due condizioni è stato condotto su dati aggregati a livello di partecipante, appaiati per confronto within-subjects. La scelta tra test parametrico (t-test appaiato) e non parametrico (Wilcoxon signed-rank) non è stata fissata a priori per ciascuna misura, ma decisa caso per caso in base alla normalità della distribuzione dei delta tra le due condizioni, verificata con il test di Shapiro-Wilk e con ispezione visiva tramite QQ plot. In tutti i casi sono riportati l'effect size (Cohen's d per i confronti parametrici, rank-biserial correlation per quelli non parametrici) e l'intervallo di confidenza al 95%, salvo dove la presenza di valori discreti ripetuti (tie) ha richiesto un livello di confidenza lievemente diverso, esplicitato nel testo. Per le correlazioni di Spearman relative a RQ5, e per la rank-biserial correlation associata ai test di Wilcoxon, non viene riportato un intervallo di confidenza sull'effect size: a differenza del coefficiente di correlazione di Pearson o della differenza media, per queste statistiche basate sui ranghi non esiste una formula standard analoga per derivarne uno analiticamente, e gli strumenti statistici comunemente usati restituiscono in questi casi solo la stima puntuale e il p-value; dove disponibile, viene riportato invece l'intervallo di confidenza sulla stima di locazione associata (spostamento mediano, Hodges-Lehmann).

Per la retention lessicale (VKS, WRT) sono stati impiegati modelli lineari a effetti misti a livello di singolo item, con la condizione sperimentale come effetto fisso di interesse e la proficiency linguistica come covariata; partecipante e item sono stati inclusi come effetti random. Le stime immediate e ritardate sono state condotte separatamente. La proficiency è stata operazionalizzata come punteggio combinato sulle bande di frequenza 4.000 e 5.000 del Vocabulary Levels Test, escludendo la banda a 3.000 per varianza pressoché nulla nel campione. La categoria linguistica dell'item (vocabolo, espressione idiomatica, ecc.) non è stata inclusa come ulteriore effetto fisso: con soli 24 item, il fattore categoria risulta quasi interamente annidato nell'identità dell'item, già trattata come effetto random, rendendo l'inclusione congiunta dei due fattori instabile.

In un caso (WRT ritardato) il modello con la struttura random completa è risultato singolare; la componente random responsabile è stata rimossa e il modello semplificato è stato adottato come riferimento primario, dopo aver verificato che il

coefficiente di interesse restasse sostanzialmente invariato tra le due specificazioni. Questa scelta è discussa nel dettaglio in §7.2.2.

Per l'analisi dell'engagement (RQ3), il confronto tra condizioni presenta un'asimmetria strutturale: le interazioni in condizione ReadLens sono ricostruibili a livello di singola porzione di testo, mentre quelle in condizione Baseline sono disponibili solo aggregate per partecipante, perché il Baseline non condivide la stessa struttura a livelli di ReadLens. Il confronto principale sul tasso di approfondimento è stato condotto sulla stessa scala (proporzione di interazioni approfondite sul totale) nonostante questa differenza strutturale, dichiarata esplicitamente come limite in §7.4.

Per l'analisi esplorativa della relazione tra profondità di esplorazione di un item e la sua retention ritardata (RQ5), è stata utilizzata la correlazione di Spearman, non parametrica, coerente con la natura ordinale della profondità di esplorazione (nessuna interazione, primo livello, secondo livello). Un controllo di robustezza ha verificato che i risultati non dipendessero dall'assunzione implicita di indipendenza tra osservazioni annidate per partecipante e per item, ripetendo la stima con un modello a effetti misti e con un'aggregazione a livello di partecipante; il dettaglio è riportato in §7.2.2.

Per tutte le analisi, il livello di significatività è stato fissato a $\alpha = .05$, con test a due code. Il campione è di 18 partecipanti, salvo dove diversamente indicato per l'esclusione di casi specifici, motivata caso per caso nel testo.

Le interviste semi-strutturate sono state analizzate tramite codifica tematica. Lo schema di codifica, inizialmente definito lungo sei dimensioni di design stabilite a priori, è stato rivisto empiricamente sulla base di un'analisi pilota: tre delle sei dimensioni originarie sono risultate trascurabili nel materiale raccolto, mentre una quota consistente dei turni conteneva contenuto valutativo rilevante non riconducibile a nessuna di esse. Le dimensioni con riscontro trascurabile sono state quindi sostituite da un insieme di categorie derivate induttivamente dal materiale stesso, mantenute solo se osservate in almeno tre partecipanti distinti. La verifica dell'affidabilità della codifica, tramite doppia codifica indipendente e calcolo di Cohen's $\varkappa$ su un campione del materiale, è stata applicata allo schema rivisto. I temi ricorrenti sono riportati in §7.2.4 e discussi in relazione ai risultati quantitativi in §7.3.

## 7.2 Risultati

Le sottosezioni seguenti presentano i risultati per ciascuna domanda di ricerca; una sintesi complessiva è riportata in Tabella 7.2

### 7.2.1 Risultati della comprensione (RQ1)

Il confronto tra le due condizioni sul punteggio di comprensione (MCQ) non mostra una differenza statisticamente significativa (Wilcoxon signed-rank appaiato, V = 59.5, p = .374): la pseudo-mediana della differenza è praticamente nulla (IC 98.1% [-0.167, 0.083]), con intervallo di confidenza che include ampiamente lo zero, e l'effect size (r = .220) è piccolo. Il punteggio medio è di 0.92 in condizione Baseline e 0.88 in condizione ReadLens, su una scala normalizzata 0-1.

Il risultato va letto insieme a un effetto soffitto marcato in entrambe le condizioni: 11 partecipanti su 18 (61%) hanno ottenuto il punteggio massimo in Baseline, e 8 su 18 (44%) in ReadLens, con deviazioni standard molto contenute in entrambi i casi (rispettivamente 0.71 e 0.75 su una scala 0-6). Questo effetto soffitto, più marcato in Baseline, comprime la variabilità residua disponibile per il confronto e riduce la sensibilità del test in entrambe le direzioni: il risultato non permette di concludere né un vantaggio né uno svantaggio del ReadLens sulla comprensione, ma riflette anche un limite dello strumento MCQ nel discriminare tra condizioni entrambe già performanti.

Un controllo di robustezza sul fattore "passaggio letto" (textA vs textB), necessario per verificare che il disegno a quadrato latino non introduca un confondimento tra testo e condizione, non rileva alcuna differenza sostanziale tra i due testi (t di Student appaiato, t(17) = -0.89, p = .386, effect size piccolo), confermando che il risultato principale non è distorto da un possibile squilibrio di difficoltà tra i due passaggi.

### 7.2.2 Risultati dell'apprendimento (RQ2-RQ5)

**Ritenzione immediata** Nessuno dei due modelli a effetti misti raggiunge la significatività statistica per l'effetto della condizione (VKS gain: b = 0.051, IC 95% [-0.074, 0.175], p = .423; WRT: b=0.143, OR = 1.15, IC 95% [0.687, 1.940], p = .588), ma in entrambi i casi il coefficiente è nella direzione prevista da RQ2 (ReadLens > Baseline). Gli intervalli di confidenza sono ampi e includono lo zero in entrambe le misure. La grandezza dell'effetto è modesta: per il VKS gain, standardizzato sulla deviazione standard osservata, corrisponde a un effect size trascurabile (d = 0.070, IC 95% [-0.101, 0.240]; $R^2$ marginale = .046, $R^2$ condizionale = .192); per il WRT, l'odds ratio di 1.15 indica un incremento modesto e incerto delle probabilità di risposta corretta nella condizione ReadLens, con un intervallo che va da un lieve svantaggio a un quasi raddoppio ($R^2$ marginale = .040, $R^2$ condizionale = .352). Con un campione di 18 partecipanti, lo studio non ha la potenza per confermare con certezza un effetto di questa entità, se presente.

La proficiency linguistica (misurata come covariata) è significativa in entrambi i modelli, con segno opposto nelle due misure: chi parte con una conoscenza del

vocabolario target già più alta guadagna in media meno punti VKS (coerente con un effetto soffitto sulla scala, confermato dalla forte correlazione negativa tra punteggio pre-test e guadagno osservato), mentre riconosce più correttamente le parole target al WRT, che non è un punteggio di guadagno e non risente dello stesso meccanismo.

**Ritenzione ritardata (7-10 giorni)** Il quadro è sostanzialmente in linea con quello immediato per il VKS gain, che resta debolmente nella direzione prevista (b = 0.037, IC 95% [-0.085, 0.159], p = .552, effect size trascurabile, d = 0.051, IC 95% [-0.116, 0.217]; $R^2$ marginale = .057, $R^2$ condizionale = .229). Per il WRT invece, il coefficiente della condizione cambia di segno rispetto alla retention immediata (b = -0.327, OR = 0.72, IC 95% [0.426, 1.219], p = .222; $R^2$ marginale = .050, $R^2$ condizionale = .261), indicando una probabilità di riconoscimento corretto leggermente più bassa nella condizione ReadLens a distanza di una settimana, con un intervallo di confidenza che include sia l'assenza di un effetto sia un moderato vantaggio o svantaggio in entrambe le direzioni, e nessuno dei due coefficienti raggiunge comunque la significatività statistica..

Il modello WRT ritardato con la struttura random completa è risultato singolare, con la componente random per partecipante collassata quasi a zero; il modello è stato quindi semplificato rimuovendo tale componente, e il coefficiente della condizione è risultato sostanzialmente identico tra le due specificazioni, confermando che la singolarità non ne comprometteva la stima.

La proficiency linguistica conferma lo stesso pattern osservato nella retention immediata, con un effetto soffitto sul VKS gain ritardato ancora più marcato. Un controllo di sensibilità che esclude i tre partecipanti con anomalie note nella somministrazione del test ritardato mostra un coefficiente quasi nullo per il VKS (b = 0.0016, contro b = 0.037 nel campione completo), ma la differenza tra le due stime è più piccola di un singolo errore standard, compatibile con normale variabilità campionaria piuttosto che con un effetto sostanziale. La conclusione di non significatività resta invariata in entrambi i casi.

**Sintesi.** Su nessuna delle due misure di retention, immediata o ritardata, RQ2 trova conferma statistica in questo campione. La direzione dell'effetto resta coerente con l'ipotesi per il VKS gain in entrambi i momenti di misurazione, mentre per il WRT il vantaggio osservato subito dopo l'esposizione non solo non persiste a distanza di una settimana, ma si inverte di segno: un'asimmetria non significativa in nessuno dei due casi, ma qualitativamente il pattern più interessante di questa sezione, ripreso in Discussione (§7.3).

**Profondità di esplorazione e retention ritardata.** Per rispondere a RQ5 si è verificato se la profondità massima di disclosure raggiunta su un item correlasse

con il relativo guadagno ritardato. RQ5 è stata dichiarata esplorativa fin dalla definizione dello studio, a differenza delle domande di ricerca primarie (RQ1-RQ4), e i suoi risultati vanno quindi interpretati con un grado di cautela maggiore rispetto ai test di conferma.

Per ciascuna delle 216 combinazioni partecipante×item osservabili in condizione Tool (18 partecipanti × 12 item), è stata calcolata la profondità massima raggiunta: nessuna interazione con l'item (126 casi), un'interazione limitata al primo livello di supporto (71 casi), o l'apertura del pannello di approfondimento (19 casi). Una sola selezione è risultata ambigua, associabile a due item contemporaneamente, ed è stata accreditata a entrambi.

La profondità di esplorazione correla positivamente con il guadagno VKS ritardato (correlazione di Spearman, rho = .337, $p < .001$, N = 216): una maggiore profondità di esplorazione di un item è associata a un guadagno maggiore nel test di retention a distanza. Non emerge invece alcuna correlazione con il punteggio WRT ritardato (rho = -.048, p = .487).

Il test di Spearman, scelto in fase di pianificazione per la natura ordinale della profondità di esplorazione, tratta implicitamente come indipendenti osservazioni che sono in realtà annidate per partecipante e per item, la stessa struttura dati trattata esplicitamente con modelli a effetti misti in §7.2.2 per RQ2. Per verificare che questa scelta non distorcesse il risultato, la stima è stata ripetuta trattando l'annidamento in due modi alternativi: un modello a effetti misti con la stessa struttura random usata per RQ2, e un'aggregazione a livello di partecipante che elimina l'annidamento. Entrambi i controlli confermano il risultato ottenuto con Spearman: la relazione positiva per il VKS gain regge in tutti e tre gli approcci, risultando anzi più marcata nell'aggregazione per partecipante (rho = .712, p = .001); l'assenza di relazione per il WRT è confermata altrettanto stabilmente, con stime vicine allo zero in tutti e tre i casi.

Questa analisi presenta un limite importante: il campione è fortemente sbilanciato verso la profondità 0, poiché un partecipante può aver appreso un item anche senza mai selezionarlo tramite lo strumento; le combinazioni a profondità 2, in particolare, coprono un numero ridotto di casi, coerente con le sole 55 aperture del pannello osservate nell'intero dataset ReadLens (il numero di volte in cui il pannello è stato aperto, distinto dal numero di singole domande poste una volta aperto il pannello, discusso in §7.2.3). Un risultato nullo su questo tipo di campione, come nel caso del WRT, non va quindi letto come evidenza di assenza di relazione, ma come limite di potenza.

Malgrado questa cautela, la correlazione osservata sul VKS gain è coerente con RQ5 e con l'ipotesi che un maggiore investimento nell'elaborazione di un termine, reso possibile dalla struttura a livelli progressivi dello strumento, sia associato a una ritenzione più solida nel tempo. Si tratta quindi di un risultato preliminare che merita approfondimento in studi con potenza adeguata.

### 7.2.3 Risultati di engagement e costo percepito (RQ3-RQ4)

Le misure raccolte per RQ3 e RQ4 osservano lo stesso fenomeno da due angolazioni diverse: da un lato il comportamento effettivo di approfondimento durante la lettura e una stima del costo dell'interazione richiesta all'utente (RQ3); dall'altro il carico di lavoro e l'usabilità percepiti, riportati direttamente dai partecipanti al termine di ciascuna condizione (RQ4). I risultati delle due prospettive sono presentati insieme perché toccano lo stesso costrutto generale, pur restando misure indipendenti l'una dall'altra.

**Dati comportamentali.** Il tasso di approfondimento (proporzione di interazioni che raggiungono il secondo livello di supporto sul totale delle interazioni) è più alto in condizione ReadLens rispetto al Baseline (differenza media di 11.1 punti percentuali, IC 95% [2.4, 19.7], $t(16) = 2.72$, $p = .015$, Cohen's d appaiato = 0.659, IC 95% [0.124, 1.177]). Questo confronto non era specificato nel piano di analisi originale, che prevedeva per l'engagement solo statistiche descrittive e la correlazione esplorativa ripresa in §7.2.2; è quindi un'estensione introdotta in fase di analisi, dichiarata come tale. Va inoltre letto tenendo conto di una differenza strutturale tra le due condizioni: ReadLens organizza il supporto in una gerarchia esplicita di livelli, mentre il Baseline mette a disposizione due strumenti indipendenti tra loro, un dizionario integrato al testo e un assistente conversazionale a lato; il tasso di approfondimento nel Baseline è quindi un costrutto ricavato a posteriori dal solo uso dell'assistente conversazionale, non la stessa misura di progressione osservata in ReadLens. Le due condizioni sono espresse sulla stessa scala ma non misurano un fenomeno strutturalmente identico, e il confronto va interpretato con questa cautela. Un partecipante è stato escluso per assenza di interazioni in almeno una delle due condizioni. Un modello a effetti misti sulla sola condizione ReadLens, che stima la probabilità di raggiungere il secondo livello dato il primo, converge senza problemi e indica una probabilità media attorno al 21% per un partecipante tipico, con circa il 18% della varianza attribuibile a differenze sistematiche tra partecipanti: l'approfondimento resta quindi un comportamento minoritario anche quando lo strumento lo rende più accessibile.

Il volume di domande poste liberamente all'IA è nettamente inferiore in condizione Baseline: solo 6 partecipanti su 18 hanno posto almeno una domanda, per un totale di 16 eventi, contro le 59 interazioni di approfondimento osservate in condizione ReadLens. Il confronto tra le due condizioni indica una differenza significativa a favore di ReadLens(Wilcoxon signed-rank appaiato, $V = 131$, $p = .027$; spostamento mediano di 1.5 domande, IC 95.5% [0, 4.5]), sebbene l'effect size ($r = .524$, formalmente grande) sia in parte trainato dal numero elevato di partecipanti Baseline senza alcuna domanda posta, e vada quindi interpretato con

questa cautela. Va inoltre segnalato che la totalità delle interazioni di approfondimento in condizione ReadLens nasce dalla selezione di un prompt suggerito, senza alcun ricorso alla domanda libera: lo strumento sposta quindi il comportamento dell'utente dalla formulazione di una domanda al riconoscimento e alla selezione di un'opzione proposta.

Le due condizioni differiscono anche nella tipologia delle domande poste. In condizione ReadLens prevale nettamente la categoria vocabolario (49% delle interazioni), seguita da riferimenti culturali (19%) ed espressioni idiomatiche (17%); il repertorio di ciascun partecipante è tipicamente concentrato su una sola categoria. In condizione Baseline, dove la codifica è stata condotta manualmente da un singolo valutatore sul numero ridotto di eventi (N = 16, senza verifica di affidabilità inter-rater), la distribuzione appare diversa: le domande relative alla comprensione di un intero passaggio, piuttosto che a singoli termini, costituiscono una quota nettamente più alta (38% contro il 7% osservato in ReadLens).

Un controllo sulla classificazione automatica mostra che tutte le domande in condizione ReadLens sono state inviate con la categoria assegnata dal sistema, senza modifiche manuali sostanziali da parte dell'utente; l'implicazione di questo dato per la fiducia riposta dagli utenti nella classificazione automatica è ripresa in Discussione (§7.3).

Infine, una stima teorica del costo di interazione (Keystroke-Level Model) indica un costo circa 4-7 volte inferiore per avviare un approfondimento in condizione ReadLens rispetto alla digitazione di una domanda libera in Baseline.

**Carico di lavoro percepito e usabilità.** Accanto a questi dati comportamentali, lo studio raccoglie anche la percezione soggettiva dei partecipanti sul carico di lavoro e sull'usabilità delle due condizioni. Il carico di lavoro percepito, misurato con il Nasa-TLX, non mostra una differenza statisticamente significativa (t-test appaiato, $t(17) = -1.73$, $p = .102$): la differenza media osservata è di -3.1 punti (IC 95% [-6.88, 0.68]) su una scala 0-100, con ReadLens leggermente più basso, effect size piccolo-medio (Cohen's d appaiato = -0.41, IC 95% [-0.88, 0.08]).

L'usabilità percepita, misurata con il SUS, non raggiunge la significatività statistica (Wilcoxon signed-rank appaiato, $V = 117$, $p = .091$). Lo spostamento mediano osservato (Hodges-Lehmann) è di +2.5 punti su scala 0-100 a favore di ReadLens (IC 96.3% [0, 6.25]). L'effect size ($r = .408$) è più ampio di quello osservato per altre misure di questo studio, ma con un campione di 18 partecipanti le stime di effect size hanno margini di incertezza ampi, e questo valore va letto con la stessa cautela di ogni altro risultato non significativo in questo studio.

Il punteggio SUS medio in condizione Baseline è già alto in termini assoluti (90.8 su 100): gli strumenti già a disposizione nella condizione di controllo sono percepiti come molto usabili di per sé, il che lascia un margine ridotto per osservare un vantaggio aggiuntivo di ReadLens nel confronto diretto. In termini assoluti,

indipendentemente dal confronto con il Baseline, ReadLens ottiene un punteggio SUS medio di 94.3 (sd = 5.9), sopra la soglia di riferimento di circa 68 comunemente citata in letteratura, e un punteggio Nasa-TLX medio di 13.1 (sd = 13.8), ben al di sotto del midpoint teorico della scala.

**Sintesi.** L'evidenza comportamentale mostra un vantaggio statisticamente significativo di ReadLens nel favorire l'engagement con il testo a un costo di interazione stimato più basso, sebbene il confronto sul tasso di approfondimento non fosse previsto dal piano di analisi originale e vada interpretato tenendo conto della diversa struttura dei due strumenti confrontati. A questo risultato comportamentale si accompagna una tendenza coerente, ma non significativa, nelle misure soggettive di carico e usabilità percepiti: con N=18, questi due risultati non possono essere interpretati come conferma dell'ipotesi, pur essendo compatibili con essa. Il dato più solido sul fronte percettivo resta la valutazione assoluta di ReadLens, che ottiene punteggi ampiamente sopra la soglia di buona usabilità e ampiamente sotto un carico percepito elevato, indipendentemente dal confronto diretto con il Baseline.

| RQ | Misura | Statistica | p | Effect size | IC 95% |
|---|---|---|---|---|---|
| RQ1 | Comprensione (MCQ) | Wilcoxon, $V = 59.5$ | .374 | $r = .220$ | $[-0.167,\ 0.083]$[a,d] |
| RQ2 | Retention immediata (VKS gain) | $b = 0.051$ | .423 | $d = 0.070$ | $[-0.101,\ 0.240]$ |
| RQ2 | Retention immediata (WRT) | $b = 0.143$ | .588 | OR = 1.15 | $[0.687,\ 1.940]$ |
| RQ2 | Retention ritardata (VKS gain) | $b = 0.037$ | .552 | $d = 0.051$ | $[-0.116,\ 0.217]$ |
| RQ2 | Retention ritardata (WRT) | $b = -0.327$ | .222 | OR = 0.72 | $[0.426,\ 1.219]$ |
| RQ3 | Tasso di approfondimento | $t(16) = 2.72$ | **.015** | $d = 0.659$ | $[0.124,\ 1.177]$ |
| RQ3 | Volume domande libere | Wilcoxon, $V = 131$ | **.027** | $r = .524$ | $[0.0,\ 4.5]$[b,d] |
| RQ4 | Carico percepito (Nasa-TLX) | $t(17) = -1.73$ | .102 | $d = -0.41$ | $[-0.88,\ 0.08]$ |
| RQ4 | Usabilità (SUS) | Wilcoxon, $V = 117$ | .091 | $r = .408$ | $[0.0,\ 6.25]$[c,d] |
| RQ5 | Profondità × retention (VKS) | Spearman, $\rho = .337$ | **<.001** | $\rho = .337$ | —[e] |
| RQ5 | Profondità × retention (WRT) | Spearman, $\rho = -.048$ | .487 | $\rho = -.048$ | —[e] |

[a] IC al 98.1%, non 95%, per la presenza di tie nei punteggi (§7.1.6).
[b] IC al 95.5%, stessa ragione.
[c] IC al 96.3%, stessa ragione.
[d] Riferito alla stima puntuale (spostamento di locazione, Hodges-Lehmann), non alla rank-biserial correlation: quest'ultima non ammette un intervallo di confidenza analitico (§7.1.6).
[e] Non disponibile per il coefficiente di Spearman (§7.1.6).

**Tabella 7.2:** Sintesi dei risultati statistici per domanda di ricerca. Le righe in grassetto indicano risultati statisticamente significativi ($\alpha = .05$).

### 7.2.4 Risultati interviste

Delle sei dimensioni di design previste per la codifica tematica, solo tre trovano riscontro sistematico nei temi effettivamente discussi dai partecipanti; la maggior parte dei temi ricorrenti è invece emersa induttivamente dal materiale stesso.

L'esperienza di lettura è stata descritta come positiva dalla generalità dei partecipanti, che hanno apprezzato la possibilità di avere a disposizione entrambi gli strumenti durante la lettura.

**Accuratezza contestuale.** Uno degli aspetti più apprezzati di ReadLens rispetto al Baseline riguarda l'accuratezza contestuale delle traduzioni fornite, percepita come capace di offrire una comprensione più immediata rispetto alle definizioni da dizionario dello strumento di controllo. Un partecipante descrive come la traduzione letterale di un termine ("bony" come "magro") non risultasse coerente con il contesto, e come l'approfondimento tramite il pannello abbia permesso di individuare il significato corretto ("pieno di lische").

**Approfondimento graduale.** Cinque partecipanti hanno descritto il valore del secondo livello di supporto come complementare, non alternativo, alla risposta immediata: utile non solo per termini sconosciuti, ma anche per approfondire parole già note, scoprendone sfumature o usi ulteriori. Un partecipante ha descritto di essere ricorso al pannello quando la sola traduzione non gli sembrava sufficiente a spiegare un riferimento specifico del testo, ottenendo così anche un chiarimento culturale che la traduzione da sola non avrebbe fornito. Un altro ha distinto esplicitamente i due livelli per funzione: la traduzione immediata per confermare termini già noti, il pannello di approfondimento per i termini realmente nuovi.

**Possibilità di selezionare porzioni di testo più ampie.** Cinque partecipanti hanno segnalato di apprezzare la possibilità, offerta da ReadLens, di selezionare più parole insieme, per ottenere una spiegazione che rispecchiasse il significato risultante dalla combinazione dei termini piuttosto che dalla sola parola isolata.

**Facilità di accesso al supporto.** Nove partecipanti hanno commentato la facilità di accesso al supporto offerto dai due strumenti; sette di questi hanno espresso una preferenza per ReadLens, in particolare per la possibilità di risolvere un dubbio senza dover formulare esplicitamente una domanda o preoccuparsi di ottenere una risposta pertinente dall'IA. Questo aspetto è coerente con il ridotto numero di interazioni registrate con l'assistente conversazionale nella condizione Baseline (§7.2.3).

**Distrazione e interruzione.** Non emergono, nel complesso, pattern che indichino una preoccupazione diffusa per un'eccessiva intrusione da parte degli strumenti. Solo due partecipanti hanno segnalato che ReadLens introduceva un'interruzione

percepita come più sensibile rispetto al Baseline, ma descritta comunque come tollerata, se non ricercata: un partecipante l'ha motivata osservando che lo strumento "fa venire voglia di approfondire".

**Lingua.** La maggioranza dei partecipanti (14 su 18) ha apprezzato la scelta di fornire le spiegazioni in italiano, mantenendo comunque gli esempi in inglese. Due partecipanti hanno espresso una preferenza per ricevere il supporto direttamente in inglese, e altri due hanno indicato il desiderio di poter scegliere la lingua a seconda della situazione.

**Annotazioni.** Quattro partecipanti hanno apprezzato la possibilità di ritrovare facilmente le parole già cercate grazie alle note create automaticamente dal sistema.

**Preferenza complessiva.** Alla domanda su quale strumento sceglierebbero per le proprie letture quotidiane, la maggioranza dei partecipanti (12 su 18) ha indicato ReadLens. Quattro hanno indicato il Baseline, motivandolo principalmente con la maggiore velocità e completezza percepite. I restanti due partecipanti hanno espresso una preferenza condizionata al tipo di lettura, sebbene in direzioni opposte tra loro: un partecipante userebbe ReadLens per la lettura quotidiana e il Baseline per lo studio approfondito, mentre l'altro preferirebbe il Baseline per la lettura veloce e il ReadLens per la lettura approfondita.

**Difficoltà percepita dei testi.** Sei partecipanti hanno riportato di aver trovato uno dei due testi più difficile dell'altro, nonostante il tentativo, in fase di selezione dei materiali, di individuare due testi comparabili per difficoltà. In cinque di questi sei casi, il testo percepito come più difficile è stato textB. Questa percezione soggettiva non trova conferma nel controllo di robustezza condotto sui punteggi MCQ (§7.2.1), che non rileva alcuna differenza sostanziale tra i due testi; questo limite è comunque discusso in §7.4.

**Testing effect.** Diversi partecipanti hanno riferito di essersi resi conto, nel corso della lettura, che i termini incontrati sarebbero stati oggetto del test successivo, orientando di conseguenza la propria attenzione verso quei termini. Questo effetto, fin dalla progettazione dello studio, è distribuito nelle due condizioni tramite il disegno a quadrato latino, non dovrebbe introdurre un vantaggio sistematico per l'una o l'altra.

## 7.3 Discussione dei risultati

Per quanto riguarda la comprensione (RQ1), il confronto tra ReadLens e Baseline non mostra alcuna differenza tra le due condizioni, in un quadro dominato da un effetto soffitto marcato in entrambe. Il ruolo dell'MCQ come strumento di misura per questo confronto è ripreso nei limiti dello studio (§7.4).

Sull'acquisizione lessicale (RQ2), lo studio non trova conferma statistica né per la retention immediata né per quella ritardata. Considerando però l'analisi esplorativa sulla sola condizione ReadLens (RQ5), emerge una correlazione positiva e stabile tra la profondità di esplorazione raggiunta su un singolo item e il relativo guadagno ritardato: è quindi possibile che il beneficio dello strumento sulla retention non sia uniforme, ma concentrato nei casi in cui l'utente sceglie effettivamente di approfondire un termine, un comportamento che nel campione osservato resta minoritario.

Le interviste aggiungono un elemento a questo quadro: l'accuratezza contestuale delle traduzioni e la possibilità di approfondire gradualmente un termine sono tra i temi più ricorrenti nell'apprezzamento espresso per ReadLens. I partecipanti valutano positivamente la possibilità di approfondimento, ma restano comunque la minoranza rispetto a chi si limita al primo livello o alla sola lettura.

Per quanto riguarda l'engagement (RQ3), ReadLens mostra un vantaggio comportamentale chiaro: un tasso di approfondimento significativamente più alto rispetto al Baseline, un volume di domande nettamente superiore, e un costo di interazione stimato più basso. Il confronto sul tasso di approfondimento va preso con cautela, perché è un'approssimazione dato che il Baseline non ha la stessa struttura a livelli di ReadLens. Le interviste offrono una possibile spiegazione a questo risultato: la facilità di accesso al supporto è uno dei temi più ricorrenti, e diversi partecipanti collegano esplicitamente la propria preferenza per ReadLens alla possibilità di risolvere un dubbio senza dover formulare una domanda o preoccuparsi di ottenerne una risposta pertinente. Questo è coerente con il dato secondo cui la totalità delle interazioni di approfondimento in condizione ReadLens nasce dalla selezione di un prompt suggerito, mai da una domanda scritta liberamente: il vantaggio comportamentale osservato non riflette solo un supporto più accessibile in astratto, ma un cambiamento specifico nel tipo di sforzo richiesto all'utente, dal formulare al riconoscere. Il fatto che nessun utente abbia mai corretto la categoria proposta dal sistema prima di inviare una domanda suggerisce inoltre un buon livello di fiducia riposta nella classificazione automatica, sebbene lo studio non ne misuri direttamente la percezione soggettiva.

Sul carico percepito e sull'usabilità (RQ4), nessuna delle due misure raggiunge la significatività statistica, ma entrambe vanno nella direzione attesa. Le interviste offrono un quadro coerente con questo risultato: non emergono pattern diffusi di preoccupazione per un'eccessiva intrusione da parte degli strumenti, e i pochi

partecipanti che segnalano un'interruzione più sensibile in ReadLens la descrivono come tollerata, se non ricercata. Questa convergenza qualitativa non compensa l'assenza di significatività statistica, ma è coerente con la lettura che i due strumenti non differiscono in modo drammatico sul piano del carico percepito, mentre la differenza più marcata tra le due condizioni resta quella comportamentale osservata in RQ3.

In termini assoluti, ReadLens ottiene comunque un punteggio SUS ampiamente sopra la soglia di buona usabilità e un punteggio Nasa-TLX ampiamente sotto un carico percepito elevato. Ciò costituisce un risultato incoraggiante sull'esperienza d'uso dello strumento, a prescindere dal confronto diretto con il Baseline.

## 7.4 Limiti dello studio

I risultati presentati vanno interpretati tenendo conto di alcuni limiti che hanno condizionato la raccolta e l'analisi dei dati.

**Potenza del campione.** Il limite principale riguarda il numero di partecipanti. Il campione di 18 partecipanti permette di rilevare con potenza adeguata solo effetti di entità medio grande. Il numero è stato un compromesso dettato dai vincoli di tempo e di fattibilità del reclutamento nel periodo estivo, quando i canali universitari sono meno frequentati del solito. Gli effect size e gli intervalli di confidenza sono stati riportati sistematicamente per ogni analisi, e i confronti su effetti di entità minore sono stati trattati come esplorativi.

**Durata dell'esposizione e del follow-up.** Lo studio ha osservato un apprendimento incidentale che per definizione richiederebbe decine di incontri distribuiti nel tempo. La ritenzione è stata misurata immediatamente dopo la lettura e a distanza di una settimana, un follow-up in linea con lo standard minimo indicato dalla letteratura, ma non sufficiente a misurare l'acquisizione a lungo termine osservata in studi con un follow-up a diversi mesi. Anche il transfer ad altri testi o generi non è stato testato.

**Selezione delle parole target e dei testi.** Le parole target sono state selezionate sulla base delle difficoltà riscontrate da un piccolo gruppo di lettori pilota. Che una parola sia risultata difficile per almeno due di questi lettori non garantisce che fosse sconosciuta anche ai partecipanti dello studio principale, un fattore che può aver influito sui guadagni osservati al VKS e al riconoscimento al WRT. Un disegno più preciso avrebbe richiesto di individuare le parole sconosciute a ciascun partecipante singolarmente e costruire o selezionare un testo su misura, una preparazione non compatibile con i tempi a disposizione.

Inoltre il VLT, usato per verificare il livello linguistico dei partecipanti tramite la banda di frequenza 3K, si basa su una corrispondenza tra le liste BNC/COCA e i livelli CEFR approssimativa. Parole nelle bande più alte con radici latine possono risultare facilmente intuibili per parlanti italiani, mentre in modo controintuitivo parole nelle bande più basse possono risultare per loro più complesse.

Per quanto riguarda la selezione dei testi, nonostante il tentativo di selezionare due passaggi comparabili per difficoltà, sei partecipanti su diciotto hanno percepito uno dei due come più difficile dell'altro, con textB indicato come più ostico in cinque di questi sei casi. Questa percezione soggettiva non trova conferma nel controllo quantitativo sui punteggi MCQ, che non rileva alcuna differenza sostanziale tra i due testi. Trovare due testi autentici realmente equivalenti in difficoltà resta comunque complesso, anche a parità di livello di competenza dichiarato, per l'influenza di interessi e ambiti di conoscenza individuali difficili da prevedere.

**Strumenti di misura.** Per il WRT non è stato misurato un guadagno, a differenza del VKS: la somministrazione è avvenuta solo dopo la lettura e nel test ritardato, senza un pre-test. Questa scelta è stata presa consapevolmente per contenere la durata della prima sessione, già di circa 75 minuti per partecipante.

L'MCQ è stato costruito senza il contributo di esperti linguistici, il cui apporto avrebbe potuto ridurre l'effetto soffitto osservato nei risultati, riconducibile a domande probabilmente troppo semplici. La natura del test non permette di distinguere se il ceiling dipenda dalla facilità delle domande, dalla facilità dei testi per il livello dei partecipanti, o da un contributo reale degli strumenti alla comprensione; un gruppo di controllo che leggesse senza alcun supporto avrebbe permesso di chiarire questo punto. Il numero contenuto di domande, sei per passaggio invece delle molte di più inizialmente previste, ha inoltre ridotto ulteriormente la potenza del confronto; anche questa riduzione è stata decisa per contenere la durata della sessione.

Infine, un campione casuale del 20% delle risposte generate da ciascuno dei due sistemi è stato valutato da un rater bilingue secondo una rubrica a tre livelli, corretta e pertinente al contesto, parzialmente corretta o debolmente contestualizzata, oppure errata o fuorviante. Il tasso di correttezza complessivo, aggregando le risposte di ReadLens e le tre disponibili in condizione Baseline, è dell'87.0% (47 su 54). Scomponendo per livello di supporto, la risposta immediata di primo livello di ReadLens raggiunge un tasso di correttezza dell'82.5% (33 su 40), sotto la soglia dell'85%, fissata in fase di pianificazione dello studio come limite minimo di accettabilità; il pannello di approfondimento di secondo livello raggiunge invece il 100% di correttezza, su un campione ridotto di 11 risposte.

**Minacce alla validità interna.** La condivisione dello schermo durante la prima sessione ha funzionato da deterrente e strumento di supervisione contro il ricorso a

risorse esterne, senza però costituire una garanzia assoluta. La seconda sessione, svolta invece da remoto per non scoraggiare la partecipazione con una seconda videochiamata, ha permesso al sistema di segnalare tempi di compilazione anomali o cambi di schermata, un controllo che non consente comunque di escludere con certezza l'uso di risorse esterne, né di confermarlo in loro assenza. Sulla base di questi segnali, tre partecipanti sono stati identificati come potenzialmente compromessi nella sessione ritardata; un controllo di sensibilità che li esclude dall'analisi del VKS è riportato in §7.2.2.

C'è da considerare poi che lo studio presenta un duplice testing effect. Da un lato, la somministrazione ripetuta dei test stessi, prima della lettura, subito dopo e a distanza di una settimana, espone di per sé i partecipanti ai termini target, costituendo una forma di esposizione ripetuta che può generare guadagni non attribuibili agli strumenti; anche se parzialmente limitata dal fatto che nessuna sovrapposizione si è verificata tra gli item del VLT e i target dello studio. Dall'altro, avendo svolto i test di comprensione e lessicali dopo la prima lettura, i partecipanti possono essersi aspettati un compito analogo anche dopo la seconda, modificando il proprio comportamento, ad esempio ricorrendo maggiormente agli strumenti disponibili durante la lettura del secondo testo. Il disegno a quadrato latino distribuisce entrambi questi effetti equamente tra le due condizioni, per cui il confronto tra ReadLens e Baseline resta valido, ma i guadagni assoluti osservati vanno considerati come potenzialmente sovrastimati rispetto a un contesto di lettura realmente incidentale. Il framing incidentale è pienamente garantito solo per il test ritardato, l'unico che i partecipanti non si aspettavano affatto. Il claim principale dello studio sull'apprendimento incidentale poggia quindi soprattutto sul dato ritardato.

**Scope deliberato dello studio.** Il setting controllato in cui si è svolta la lettura non equivale a una lettura naturalistica. Per mitigare questa distanza sono stati usati testi autentici, con lettura libera e istruzioni non direttive sull'uso degli strumenti.

I tre livelli della gerarchia di supporto sono stati valutati nel loro insieme: il contributo del singolo livello alla comprensione o alla ritenzione non è isolabile con questo disegno.

Infine, le misure di apprendimento utilizzate coprono solo il vocabolario, inclusi idiomi ed espressioni. Riferimenti culturali e forme grammaticali non sono stati inclusi nei test di ritenzione, che per costruzione misurano solo conoscenza lessicale. I claim relativi all'apprendimento restano quindi circoscritti alla dimensione lessicale.

# Capitolo 8

# Conclusioni

Questo lavoro di tesi ha esplorato come il supporto alla lettura in L2 possa essere ripensato integrando l'intelligenza artificiale generativa direttamente nel flusso della lettura, senza costringere il lettore a interromperla per consultare risorse esterne. A partire dallo scarto identificato tra il quadro teorico sulla relazione tra lettura e acquisizione lessicale in L2 e le scelte concrete di interazione, è stato condotto uno studio formativo con lettori italiani di inglese e esperte di lingua, i cui risultati hanno guidato la definizione dei requisiti e le scelte di design del prototipo.

Il risultato di questo percorso è Readlens, un'applicazione web che offre traduzioni e spiegazioni contestuali di parole ed espressioni direttamente all'interno del testo, su due livelli di approfondimento: una risposta immediata data da una traduzione contestualmente accurata, e un pannello di approfondimento. Lo strumento è stato realizzato seguendo un'architettura client-server, con un backend in Node.js e TypeScript con database PostgreSQL, e un frontend in React. È stato poi valutato attraverso uno studio controllato con 18 partecipanti, che ne ha confrontato il contributo alla comprensione e all'apprendimento del lessico, l'engagement generato durante la lettura, e il carico di lavoro e l'usabilità percepiti, rispetto a una soluzione di riferimento basata su strumenti generici già esistenti.

Lo studio non ha trovato una differenza statisticamente significativa tra Readlens e la soluzione di riferimento né sulla comprensione del testo né sulla ritenzione del vocabolario, immediata o a distanza di una settimana. Il dato più solido riguarda invece l'engagement: Readlens genera un tasso di approfondimento significativamente più alto, sostenuto da un costo di interazione più basso rispetto alla formulazione di una domanda libera. Il carico di lavoro percepito e l'usabilità non mostrano differenze significative rispetto alla soluzione di riferimento, ma Readlens ottiene in termini assoluti punteggi ampiamente positivi su entrambe le misure.

## 8.1 Limitazioni e lavoro futuro

I risultati ottenuti vanno interpretati alla luce di alcuni limiti, in parte legati agli studi condotti e in parte allo strumento stesso.

Sia lo studio formativo sia lo studio controllato hanno coinvolto un numero ridotto di partecipanti, reclutati con campionamento di convenienza in ambito universitario. Questo limita la potenza statistica a disposizione per rilevare eventuali differenze tra le condizioni: un campione più ampio permetterebbe di verificare con maggiore precisione se le stime osservate, per quanto non significative in questo studio, riflettano un effetto reale di piccola entità. Il campione risulta inoltre ristretto nella fascia d'età rappresentata, prevalentemente giovane e universitaria, il che limita la possibilità di generalizzare i risultati ad altre popolazioni di apprendenti. Un ampliamento del reclutamento, anche al di fuori del contesto universitario, resta comunque un passo naturale per uno studio futuro.

Un ulteriore spunto emerge dai dati sulla proficiency linguistica. I partecipanti con un livello di inglese più alto hanno mostrato guadagni VKS minori, un pattern coerente con un effetto soffitto. Il campione di questo studio, per coerenza con la popolazione dello studio formativo, era composto da apprendenti di livello B2-C1; un'ipotesi interessante, seppur speculativa, è che uno strumento come Readlens possa offrire un contributo più marcato per apprendenti di livello più basso, non ancora testati.

Indipendentemente dalla dimensione del campione, lo studio ha osservato l'effetto di una singola sessione di lettura. L'acquisizione lessicale stabile richiede invece incontri ripetuti con una parola distribuiti nel tempo: uno studio di deployment più esteso permetterebbe di osservare l'effetto di Readlens sulla ritenzione lessicale nel lungo periodo, cosa che il disegno attuale non può cogliere. Resta inoltre da chiarire il motivo per cui la stima sul WRT cambia segno tra la misurazione immediata e quella a distanza di una settimana, pur restando entrambe non significative: un follow-up più lungo, con più punti di misurazione intermedi, potrebbe aiutare a distinguere un pattern reale da semplice rumore campionario.

La preparazione dei materiali e degli strumenti di misura beneficerebbe infine del contributo di esperti del settore SLA, pur essendo stati costruiti seguendo criteri e strumenti tratti dalla letteratura.

Rimane inoltre interessante la correlazione osservata, all'interno della sola condizione Readlens, tra profondità di esplorazione e guadagno lessicale ritardato. Se un maggiore investimento nell'elaborazione di un termine è associato a una ritenzione più solida, come questo risultato esplorativo suggerisce, uno strumento come Readlens potrebbe offrire un contributo più marcato in un contesto di apprendimento intenzionale, dove l'approfondimento sia esplicitamente richiesto o incoraggiato, rispetto al contesto incidentale qui indagato, in cui la scelta di approfondire resta minoritaria e lasciata all'utente.

Dal punto di vista delle funzionalità dello strumento, si possono considerare per sviluppi futuri alcune delle funzionalità rimandate per non introdurre variabili aggiuntive nello studio, ma riconosciute come potenzialmente utili all'apprendimento. Alcuni partecipanti hanno espresso il desiderio di ricevere spiegazioni anche nella lingua target: si potrebbe quindi reintrodurre la possibilità di scegliere la lingua delle risposte, eventualmente accompagnata da un adattamento del livello di inglese usato al livello del lettore, sfruttando le capacità di personalizzazione dell'IA. Resta inoltre aperto un interrogativo su come reintrodurre, in una forma diversa, attività che richiedano una produzione attiva da parte dell'utente, a cui si è consapevolmente rinunciato nel prototipo valutato.

Infine, va migliorata la qualità delle risposte del primo livello dello strumento, il cui tasso di correttezza è risultato leggermente inferiore alla soglia inizialmente definita come accettabile. L'intervento è tanto più importante considerando che il primo livello, secondo quanto emerso dalle interviste, è tra gli aspetti più apprezzati dello strumento.

Dal punto di vista tecnico, un limite riguarda la mancata pulizia, a livello di database, dello storico delle conversazioni quando un'annotazione viene cancellata. Altre direzioni di sviluppo tecnico riguardano più da vicino le prestazioni e la piattaforma dello strumento: l'addestramento o l'adattamento di un modello linguistico più piccolo, eseguibile in locale e specializzato sul compito di spiegazione contestuale, potrebbe ridurre il costo e la latenza per richiesta rispetto alla dipendenza attuale da un modello esterno tramite API. Sul piano della piattaforma, un'estensione browser resterebbe l'evoluzione più naturale dello strumento, permettendo di intercettare la lettura direttamente nel contesto in cui avviene, indipendentemente dalla fonte del testo, senza richiedere all'utente di caricare manualmente ciò che sta leggendo. Un adattamento a dispositivi mobili, come tablet o smartphone, rimane un'ulteriore direzione da esplorare.

# Bibliografia


[1] R. Y. Pang, H. Schroeder, K. S. Smith, S. Barocas, Z. Xiao, E. Tseng e D. Bragg. «Understanding the LLM-ification of CHI: Unpacking the Impact of LLMs at CHI through a Systematic Literature Review». In: *Proc. 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. ACM, 2025. DOI: 10.1145/3706598.3713726 (cit. a p. 1).

[2] G. L. Liu e X. Zhao. «A Scoping Review of AI-Mediated Informal Language Learning: Mapping Out the Terrain and Identifying Future Directions». In: *ReCALL* 38.1 (2026), pp. 111–130. DOI: 10.1017/S0958344025100359 (cit. a p. 1).

[3] S. Chung, H. Jeon, S. Shin e M. N. Hoque. *Reading.help: Supporting EFL Readers with Proactive and On-Demand Explanation of English Grammar and Semantics*. arXiv:2505.14031v2. 2025. DOI: 10.48550/arXiv.2505.14031 (cit. alle pp. 2, 11).

[4] Y. Yang, S. W. Lee, J. Y. Song, S. Yun e Y.-H. Kim. «LingoQ: Bridging the Gap between EFL Learning and Work through AI-Generated Work-Related Quizzes». In: *Proc. 2026 CHI Conference on Human Factors in Computing Systems (CHI '26)*. ACM, 2026, pp. 1–23. DOI: 10.1145/3772318.3791342 (cit. a p. 2).

[5] A. Trusty e K. N. Truong. «Augmenting the Web for Second Language Vocabulary Learning». In: *Proc. SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. New York, NY, USA: ACM, 2011, pp. 3179–3188. DOI: 10.1145/1978942.1979414 (cit. a p. 2).

[6] S. Khezrlou, R. Ellis e K. Sadeghi. «Effects of Computer-Assisted Glosses on EFL Learners' Vocabulary Acquisition and Reading Comprehension in Three Learning Conditions». In: *System* 65 (2017), pp. 104–116. DOI: 10.1016/j.system.2017.01.009 (cit. alle pp. 2, 6, 75, 77).

[7] A. Cooper, R. Reimann, D. Cronin e C. Noessel. *About Face: The Essentials of Interaction Design*. Fourth. Indianapolis, IN: John Wiley & Sons, 2014. ISBN: 978-1-118-76657-6 (cit. alle pp. 3, 31).

[8] W. Grabe e J. Yamashita. *Reading in a Second Language: Moving from Theory to Practice*. Second. The Cambridge Applied Linguistics Series. Cambridge, UK: Cambridge University Press, 2022 (cit. alle pp. 5, 6, 8).

[9] J. H. Hulstijn. «Incidental and Intentional Learning». In: *The Handbook of Second Language Acquisition*. A cura di C. J. Doughty e M. H. Long. Malden, MA: Blackwell Publishing, 2003, pp. 349–381 (cit. a p. 5).

[10] P. Ur. «Re-examining Some Conventional Assumptions in Vocabulary Teaching: What Can We Learn from the Research?» In: *The Handbook of Practical Second Language Teaching and Learning*. A cura di E. Hinkel. New York, NY: Routledge, 2022, pp. 409–420 (cit. alle pp. 5, 6, 8–10).

[11] R. Waring e M. Takaki. «At What Rate Do Learners Learn and Retain New Vocabulary from Reading a Graded Reader?» In: *Reading in a Foreign Language* 15.2 (2003), pp. 130–163 (cit. alle pp. 5, 8).

[12] M. Pigada e N. Schmitt. «Vocabulary Acquisition from Extensive Reading: A Case Study». In: *Reading in a Foreign Language* 18.1 (2006), pp. 1–28 (cit. a p. 5).

[13] M. F. Teng. «Understanding Incidental Vocabulary Learning in Practice». In: *Asian Journal of English Language Teaching* 32.1 (2023), pp. 7–28 (cit. a p. 6).

[14] I. S. P. Nation. *Learning Vocabulary in Another Language*. Third. Cambridge, UK: Cambridge University Press, 2022. ISBN: 978-1-009-09835-9. DOI: 10.1017/9781009093873 (cit. alle pp. 6, 9).

[15] J. H. Hulstijn e B. Laufer. «Some Empirical Evidence for the Involvement Load Hypothesis in Vocabulary Acquisition». In: *Language Learning* 51.3 (2001), pp. 539–558. DOI: 10.1111/0023-8333.00164 (cit. a p. 6).

[16] A. Yanagisawa e S. Webb. «Involvement Load Hypothesis Plus: Creating an Improved Predictive Model of Incidental Vocabulary Learning». In: *Studies in Second Language Acquisition* 44.5 (2022), pp. 1279–1308. DOI: 10.1017/S0272263121000577 (cit. alle pp. 7, 10).

[17] S. Webb, T. Uchihara e A. Yanagisawa. «How Effective Is Second Language Incidental Vocabulary Learning? A Meta-Analysis». In: *Language Teaching* 56.2 (2023), pp. 161–180. DOI: 10.1017/S0261444822000507 (cit. a p. 8).

[18] Rob Waring. «Extensive Reading». In: *Handbook of Practical Second Language Teaching and Learning*. A cura di Eli Hinkel. Routledge, 2022 (cit. a p. 8).

[19] J. Yun. «The Effects of Hypertext Glosses on L2 Vocabulary Acquisition: A Meta-Analysis». In: *Computer Assisted Language Learning* 24.1 (2011), pp. 39–58 (cit. a p. 9).

[20] H. Lee e J. H. Lee. «The Effects of Electronic Glossing Types on Foreign Language Vocabulary Learning: Different Types of Format and Glossary Information». In: *Asia-Pacific Education Researcher* 24.4 (2015), pp. 591–601. DOI: 10.1007/s40299-014-0204-3 (cit. a p. 9).

[21] F. Faghfouri, S. H. Aliasin e E. Mohammadi. «Computer-Mediated Immediate and Delayed L1 and L2 Glosses and Vocabulary Learning and Reading Comprehension of an ESP Text». In: *Computer Assisted Language Learning Electronic Journal (CALL-EJ)* 23.1 (2022), pp. 445–465 (cit. a p. 9).

[22] N. Miyasako. «Does Text-Glossing Have Any Effects on Incidental Vocabulary Learning through Reading for Japanese Senior High School Students?» In: *Language Education & Technology* 39 (2002), pp. 1–20. DOI: 10.24539/let.39.0_1 (cit. a p. 9).

[23] I. Udry e R. Berthele. «A Case Study of Online and Paper Bilingual Dictionaries and their Impact on Vocabulary Learning through Meaning-Focused Reading». In: *International Journal of Lexicography* 38.3 (2025), pp. 219–237. DOI: 10.1093/ijl/ecaf004 (cit. a p. 10).

[24] Cheng Yuxin, Wang Ruochen, Chen Jiawei, Chao Yijun, Aliye Maimaitili e Zhang Haoruo. «Context-Based AI Translation From a Globalization Perspective: A Case Study of ChatGPT». In: *Sino-US English Teaching* 20.9 (2023). DOI: 10.17265/1539-8072/2023.09.005 (cit. alle pp. 10, 11).

[25] M. F. Lungu, L. Van Den Brand, D. Chirtoaca e M. Avagyan. «As We May Study: Towards the Web as a Personalized Language Textbook». In: *Proc. 2018 CHI Conference on Human Factors in Computing Systems*. 2018, pp. 1–12. DOI: 10.1145/3173574.3173912 (cit. a p. 10).

[26] M. Azab, A. Salama, K. Oflazer, H. Shima, J. Araki e T. Mitamura. «An NLP-Based Reading Tool for Aiding Non-Native English Readers». In: *Recent Advances in Natural Language Processing (RANLP)*. 2013, p. 41 (cit. a p. 11).

[27] S. Amershi et al. «Guidelines for Human-AI Interaction». In: *Proc. 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. New York, NY, USA: ACM, 2019, pp. 1–13. ISBN: 978-1-4503-5970-2. DOI: 10.1145/3290605.3300233 (cit. a p. 55).

[28] J. P. Kincaid, R. P. Fishburne, R. L. Rogers e B. S. Chissom. *Derivation of New Readability Formulas for Navy Enlisted Personnel*. Research Branch Report 8-75. Naval Air Station Memphis, 1975 (cit. a p. 74).

[29] T. Cobb. *Compleat Web-VP v.3.1 [computer program]*. https://www.lextutor.ca/vp/comp/. Accessed 10 Jul 2026. 2026 (cit. a p. 74).

[30] S. Webb, Y. Sasao e O. Ballance. «The Updated Vocabulary Levels Test: Developing and Validating Two New Forms of the VLT». In: *ITL – International Journal of Applied Linguistics* 168.1 (2017), pp. 33–69. ISSN: 0019-0829. DOI: 10.1075/itl.168.1.02web (cit. a p. 75).

[31] T. S. Paribakht e M. Wesche. «Vocabulary Enhancement Activities and Reading for Meaning in Second Language Vocabulary Acquisition». In: *Second Language Vocabulary Acquisition*. A cura di J. Coady e T. Huckin. First. Cambridge University Press, 1996, pp. 174–200. DOI: 10.1017/CBO9781139524643.013 (cit. a p. 75).

[32] S. G. Hart e L. E. Staveland. «Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research». In: *Human Mental Workload*. A cura di P. A. Hancock e N. Meshkati. Vol. 52. Advances in Psychology. Amsterdam: North-Holland, 1988, pp. 139–183. DOI: 10.1016/S0166-4115(08)62386-9 (cit. a p. 75).

[33] S. Borsci, S. Federici e M. Lauriola. «On the Dimensionality of the System Usability Scale: A Test of Alternative Measurement Models». In: *Cognitive Processing* 10.3 (2009), pp. 193–197. DOI: 10.1007/s10339-009-0268-9 (cit. a p. 76).

[34] J. Brooke. «SUS: A 'Quick and Dirty' Usability Scale». In: *Usability Evaluation in Industry*. A cura di P. W. Jordan, B. Thomas, B. A. Weerdmeester e I. L. McClelland. London: Taylor & Francis, 1996, pp. 189–194 (cit. a p. 76).