frontiers in Language Sciences

TYPE Original Research
PUBLISHED 28 October 2026
DOI 10.3389/flang.2026.1912839

Check for updates

OPEN ACCESS

EDITED BY
Ifigeneia Dosi,
Democritus University of Thrace, Greece

REVIEWED BY
Sangyub Kim,
Chonnam National University,
Republic of Korea
Laura Nadal,
Ca' Foscari University of Venice, Italy

*CORRESPONDENCE
Maria Vender
maria.vender@univr.it

RECEIVED 18 June 2026
REVISED 25 September 2026
ACCEPTED 28 September 2026
PUBLISHED 28 October 2026

CITATION
Vender M, Venagli I, Nardon A, Piccinin S, Dal Maso S and Melloni C (2026) Processing event nominalizations in developing readers: eye-tracking evidence from bilingual and dyslexic readers. *Front. Lang. Sci.* 5:1912839. doi: 10.3389/flang.2026.1912839

COPYRIGHT
© 2026 Vender, Venagli, Nardon, Piccinin, Dal Maso and Melloni. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Processing event nominalizations in developing readers: eye-tracking evidence from bilingual and dyslexic readers

Maria Vender*, Ilaria Venagli, Andrea Nardon, Sabrina Piccinin, Serena Dal Maso and Chiara Melloni

Department of Cultures and Civilizations, University of Verona, Verona, Italy

Introduction: Event nominalizations are a central feature of academic language: they compress complex events into nominal structures, increasing informational density while reducing the transparency of thematic-role assignment. The present eye-tracking study examined whether nominalizations impose greater processing demands than corresponding active and passive constructions in developing readers, and whether these effects vary across readers with different linguistic and literacy profiles.

Methods: Seventy-nine fifth- and sixth-grade children participated in the study: 28 typically developing monolingual Italian children, 21 monolingual Italian children with developmental dyslexia, and 30 typically developing bilingual children with migrant background and Italian as the societal language. Participants silently read short passages in which the same target event was expressed either as a nominalization, a passive construction, or an active construction, and then answered comprehension questions while their eye movements were recorded. Accuracy, standard eye-movement measures, and scanpaths were analyzed to capture comprehension outcomes and online reading strategies.

Results: Nominalizations did not significantly reduce comprehension accuracy, but they affected online processing during both initial reading and the question-answering phase. Scanpath dissimilarity was higher for nominalizations than for both verbal constructions during initial reading, and than for passives during question-answering in correctly answered trials. In these trials, nominalizations were also associated with a higher probability of partial rescanning and a lower probability of minimal rescanning relative to active texts, indicating greater reliance on rereading when answering questions correctly. Moreover, children with dyslexia showed a generalized increase in processing effort, whereas bilingual children displayed an intermediate profile, with increased reliance on rereading during question-answering.

Discussion: These findings indicate that nominalizations significantly influence online reading processes even when questions are answered correctly, suggesting additional demands in retrieving information about event participants. Overall processing patterns differed across reader groups, and children with dyslexia showed the greatest effort in both online reading and question-answering.

KEYWORDS
bilinguals with migrant background, developing readers, developmental dyslexia, event nominalizations, eye-tracking, reading comprehension

1 Introduction

Achieving efficient reading comprehension is fundamental to academic success and, more broadly, to an individual's professional and cultural development. Although text comprehension is so pervasive in contemporary societies that it may appear self-evident, it is in fact a highly sophisticated ability, resulting from the dynamic interaction of multiple processes operating at different levels, including both lower- and higher-level components (Grabe and Yamashita, 2022). Lower-level processes comprise decoding, which must become automatized to support fluent comprehension, as well as core linguistic mechanisms such as lexical access, syntactic parsing, and semantic encoding, all of which provide the foundational input on which higher-level comprehension processes can build. Higher-level processes, in turn, enable readers to move beyond sentence-level, literal meaning to construct a mental representation of the text as a whole, by integrating textual information with background knowledge, generating inferences, and monitoring comprehension. Within the Reading Systems Framework, vocabulary plays a key role, which serves as an interface between lower-level word identification processes and higher-level word-to-text integration processes (Perfetti and Stafura, 2014). In this framework, lexical representations, encompassing phonological, orthographic, semantic, and morphosyntactic information, constitute a critical pressure point in the reading system, with the consequence that the efficiency and precision of reading comprehension are strongly determined by the quality and accessibility of an individual's lexical representations (Perfetti, 2007).

Fragilities at the lexical level can therefore propagate both downward, affecting decoding and lexical access, and upward, compromising word-to-text integration and, consequently, sentence and text comprehension. Importantly, these processes unfold under working memory constraints, as comprehension depends on the activation, maintenance, and integration of lexical and propositional information within a system of limited processing resources. Accordingly, difficulties in reading comprehension may therefore arise not only from underspecified lexical representations, but also from reduced efficiency in coordinating linguistic processes under memory constraints. This is particularly relevant in populations in which linguistic representations and cognitive resources may be more fragile, as in children with dyslexia and bilingual children. Both populations indeed present weaknesses in reading comprehension that might seem superficially similar, although they arise from different underlying deficits.

Nominalizations, which are widespread in academic texts, constitute a particularly informative domain for investigating these vulnerabilities, since they increase lexical density and informational compression, requiring readers to unpack morphologically complex forms and recover implicit argument structures. For this reason, they represent a demanding interface between word-level processing and discourse-level integration, although only a few studies have systematically examined their acquisition and online processing, particularly in populations with linguistic fragilities. To fill this gap, we developed an eye-tracking study to examine how monolingual children with and without dyslexia and bilingual migrant children tested in the societal language process texts with nominalizations, as will be discussed in the following sections.

1.1 Nominalizations: characteristics and effects on reading comprehension

Nominalization is the linguistic process by which elements belonging to other word classes (typically verbs) are transformed into nouns, allowing actions, processes, or states to be encoded in a nominal configuration (Comrie and Thompson, 2007; Gaeta, 2002). Event nominalizations, in particular, are deverbal nouns that preserve the actional properties of their base verb while retaining its argument structure (Grimshaw, 1990; Koptjevskaja-Tamm, 1993; Melloni, 2011). Consider, for instance, the Italian sentences below:

(1) La Gran Bretagna abolì il commercio degli schiavi

"Great Britain abolished the slave trade"

(2) L'abolizione del commercio degli schiavi da parte della Gran Bretagna

"The abolition of the slave trade by Great Britain"

In the nominalized version in (2), the event expressed in (1) by the verb *abolì* ("abolished") is reconfigured as the head of a complex noun phrase (*abolizione*, "abolition"). Although the semantic content is preserved, the thematic roles are not mapped onto canonical positions: the internal argument is encoded as a genitive complement (*del commercio degli schiavi*, "of the slave trade"), and the external argument is expressed through a by-phrase (*da parte della Gran Bretagna*, "by Great Britain"), which is syntactically optional. This restructuring reduces the transparency of agent–patient relations and contributes to the more impersonal and abstract style typically associated with academic prose (Gaeta, 2002; Grimshaw, 1990; Halliday and Martin, 1993). Nominalizations are, in fact, typically associated with more objectified modes of representation, since, unlike finite clauses, which are anchored to tense, chronological sequencing, and explicit agency, they detach processes from their temporal anchoring, backgrounding the agent, and expressing them as abstract entities (Downing, 2000). In addition, by compressing propositional content into a compact nominal they increase the text's informational density, and support the condensed style characteristic of academic discourse. Finally, nominalizations enhance textual cohesion: an action described in a clause, indeed, can be subsequently encapsulated and reintroduced in the discourse as a nominal expression (e.g., The government approved the reform. The *approval*…), thus supporting the creation of coherent anaphoric chains and facilitating the hierarchical organization of information in a text and discourse progression (Comrie and Thompson, 2007).

These features render nominalizations a key component of written academic language, especially in science, mathematics, history, and law, with an intensification in the last 50 years across disciplines (Biber and Gray, 2016; Hyland and Jiang, 2021). Several corpus-based studies have indeed shown that nominalizations occur significantly more frequently in written academic registers than in spoken language (Biber et al., 1999) and are particularly

dense in scientific and educational texts (Fang et al., 2006). Importantly, they are already present in primary school textbooks, with percentages rising to over 5% in middle school materials and reaching similarly high levels from grade three onward (Mueller, 2015).

Nevertheless, it should be noted that while nominalizations enhance textual conciseness and cohesion, the discourse-level advantages discussed above may come at the cost of increased cognitive demands during comprehension. As a matter of fact, the same features that contribute to textual efficiency can also reduce content accessibility: the structure of the event compressed into a nominal unit needs to be unpacked by the reader in order to reconstruct the thematic relations and achieve proper understanding (Bello, 2016). As observed above, indeed, nominalizations reduce the transparency of the mapping between syntactic structure and thematic roles, requiring readers to reconstruct the underlying clausal relations to identify event participants and to integrate them into the unfolding discourse representation (Ravelli, 1996; Schleppegrell, 2004). An additional inferential passage is required when the agent is backgrounded or omitted, as the reader must retrieve implicit information in order to reconstruct the event structure. Similarly, information about tense and aspect needs to be extracted from the context, as they are not expressed by nominals, unlike their verbal counterparts. Notice, moreover, that nominalized expressions frequently carry an existential presupposition: when an event is encoded as a definite description, its existence is taken for granted rather than asserted. By presenting the event as given background information, nominalization renders this assumption implicit, which could potentially further obscure interpretative transparency while simultaneously strengthening the persuasive force of the message (Cominetti, 2023; Givón, 1984). As a result, processing might be slowed, particularly when contextual support is limited or when readers' linguistic and cognitive resources are still developing. Notice in fact that all the inferential work required to process nominalizations in texts unfold under working memory constraints, as readers must retrieve the event's structure, while integrating new incoming information, which may negatively impact text comprehension, especially for readers whose lexical representations or cognitive resources are less robust (Perfetti and Stafura, 2014).

Consistently, Halliday and Martin (1993:10) describe nominalization as a structure that is at once "freeing" and "constraining," observing that while it enables the packaging of complex processes into compact semiotic units and facilitates the organization of scientific discourse, it may simultaneously hinder comprehension. However, it should be noted that the amount of effort required to unpack nominalizations and to understand them properly is constrained by the type of nominalization and the maturity and expertise of the reader. Regarding the former, Plemenitaš (2015) observed that highly lexicalized nominalizations (e.g., experiment) may be processed through relatively automatic lexical access, whereas less entrenched or more compositionally transparent forms (e.g., plumpness) require readers to analyze the internal morphological structure of the nominal form and retrieve the underlying verbal representation. This is also related to the reader's knowledge: younger and less expert readers, as well as fragile readers, are expected to struggle more with sentences containing nominalizations (Halliday and Martin, 1993), especially when they are less frequent and less recognizable. Unlike other complex structures, such as passives, nominalizations are less transparent and might be less easily identified, due to their high formal heterogeneity, as they can result from different morphological processes. Along with nominals obtained by adding a derivational suffix to the verb base (e.g., *-zione* in *abolizione*), there are indeed cases of conversion and zero derivation (e.g., *sviluppo* "development" from *sviluppare* "develop," Iordăchioaia and Melloni, 2023), where it might be more difficult to identify the underlying verbal base, especially for less experienced readers, thus hindering the event reconstruction with possible negative effects on text comprehension as a whole.

1.2 Empirical evidence on the processing of nominalizations

Despite the identification of nominalization as a source of linguistic compression and abstraction particularly present in textbooks and academic language, systematic empirical investigations of its processing remain relatively limited. The available evidence converges in suggesting that it imposes measurable processing costs, even when comprehension accuracy is preserved. Using a self-paced reading paradigm, Hansen et al. (2006) compared adult German speakers' comprehension of original legal texts containing nominalizations with reformulated versions in which nominals were partially unpacked into verbal clauses. Results showed that nominalizations elicited longer reading times compared to semantically equivalent verbal structures and that moderate reformulation reduces processing time and response latency, whereas excessive simplification does not necessarily enhance comprehension accuracy. These findings confirm that nominalizations increase processing effort, but, on the other hand, they suggest that structural simplification alone does not automatically improve understanding, likely because cohesion and information structuring are simultaneously affected. Eye-tracking studies provide further insights into the online processing of nominalizations. In particular, Wolfer (2016) examined adults reading German legal texts with nominalizations and reformulated verbal versions, and reported that nominalizations elicit longer first-pass fixation durations, increased total reading time, and more regressions than non-nominalized alternatives, thus indicating higher processing effort. Crucially, comprehension accuracy was similar across conditions, suggesting that nominalizations do not necessarily lead to decreased accuracy, but rather impose higher processing costs that are more readily captured by online measures.

Developmental evidence suggests that the processing demands associated with nominalized constructions may be amplified in younger readers, although research is extremely limited in this respect. Yatsushiro and Alexiadou (2020) showed that German-speaking children (mean age 4.8 years old) exhibited greater difficulty than adults, who performed at ceiling, in interpreting deverbal nominalizations in *-ung* in a picture selection task. Interestingly, the authors observed that children's performance

with nominalizations parallels well-documented patterns in the acquisition of verbal passives, although they did not directly test this hypothesis. Specifically, difficulty increased when the external argument was overtly expressed via a *durch*-phrase (by-phrase), a configuration in which argument structure is less transparently aligned with surface form. Based on these findings, the authors suggest that children experience particular challenges in assigning thematic roles within nominal constructions and that computing non-canonical argument structure, which is shared by nominalization and passive structures, might be particularly demanding for children. Recent eye-tracking evidence in children further supports the processing cost associated with nominalization. In a study on arithmetic word problem solving with 10 to 13-year-old children, Daroczy et al. (2025) showed that passages with nominalizations increased solution times and total fixation durations, as compared to verbalized forms. To the best of our knowledge, no other studies directly tested children's processing of nominalizations.

1.3 Nominalizations and linguistic vulnerability: expectations and empirical gaps

While the existing empirical literature has primarily focused on skilled adult readers, the processing demands associated with nominalized constructions may be particularly pronounced for children with linguistic fragilities and text comprehension difficulties. In particular, children with dyslexia and bilinguals, especially those with migrant background, represent two interesting populations, as they present distinct profiles of literacy and linguistic vulnerability that converge at the level of reading comprehension, albeit for different underlying reasons.

Developmental dyslexia is classified within Specific Learning Disorders as a neurodevelopmental condition characterized by a marked impairment in the acquisition of reading skills that cannot be explained by low intelligence, sensory deficits, neurological disorders, or inadequate educational opportunities (First, 2013; Lyon et al., 2003). Prevalence estimates range from 5 to 17.5%, depending on orthographic transparency, with a worldwide prevalence of approximately 7% (Yang et al., 2022). A large body of research has shown that reading deficits in dyslexia are modulated by orthographic depth (Carioti et al., 2021). Poor reading fluency emerges as the core and universal deficit of dyslexia, which although modulated by orthographic depth, persists across orthographies and into adulthood, and reflects impaired phonological decoding and reduced automatization (Carioti et al., 2021; Ziegler and Goswami, 2005). Beyond decoding difficulties, dyslexia is associated with broader linguistic and cognitive vulnerabilities. Among them, phonological deficits are well documented (Snowling et al., 2020), alongside difficulties in rapid automatized naming and lexical access (Araújo et al., 2015), morphological competence (Casalis et al., 2004; Georgiou et al., 2023; Melloni and Vender, 2022), grammatical competence (Cardinaletti et al., 2022), as well as working memory and attentional control (Berninger et al., 2008; Vender, 2017).

The presence of these deficits suggests that dyslexia cannot be reduced to a purely phonological disorder but may involve multiple fragile points within the reading system, affecting linguistic competence, literacy as well as reading comprehension. A meta-analysis by Georgiou et al. (2022) shows, in fact, that children and adolescents with dyslexia display a substantial deficit in reading comprehension compared to age-matched controls, although this deficit is typically smaller than that observed for decoding. Importantly, comprehension difficulties are not fully explained by lower decoding deficits, as they persist even compared with reading-level-matched controls, and, albeit to a lower extent, in listening comprehension (Georgiou et al., 2022). Comprehension problems in dyslexia thus seem to arise from the interaction between decoding inefficiency, linguistic knowledge, and limited working memory and processing resources. From the perspective of the Reading Systems Framework, they can be conceived as the outcome of both knowledge-based fragilities and process-based inefficiencies, operating under constrained working memory resources: when lower-level processes are slow and effortful, they consume resources that would otherwise be available for higher-level integration, inference generation, and monitoring, potentially leading to increased rereading and reduced processing efficiency even when overt comprehension accuracy remains relatively high.

Linguistic vulnerabilities are often reported among bilinguals as well, especially regarding vocabulary knowledge and lexical access (Bialystok et al., 2008, 2010). Morphosyntactic competence may also be fragile, particularly in contexts of limited input quantity and quality (Marinis and Chondrogianni, 2010; Stepanov et al., 2026; Vender et al., 2018). A large-scale meta-analysis indicates that bilingual children reading in their L2 display a moderate deficit in reading comprehension too, compared to monolingual peers (Melby-Lervåg and Lervåg, 2014). This is particularly relevant in migrant backgrounds, as reported by international assessments that point to a marked difficulty in reading comprehension for bilinguals (OECD, 2023). Crucially, this gap is primarily driven by weaknesses in language comprehension (vocabulary and listening comprehension), whereas differences in decoding and phonological awareness are generally smaller (Marinelli et al., 2026; Novita et al., 2022). Importantly, difficulties are greater in children from lower socioeconomic backgrounds, and in situations where exposure to academic language is limited, leading to educational inequalities (Aguiar et al., 2020; Novita et al., 2022). Educational data from the Italian context, where migrant children constitute more than 11% of the school population, confirm stable performance gaps in reading comprehension tasks, particularly among first-generation students, with differences closely linked to Italian language proficiency and academic vocabulary (INVALSI, 2025). These difficulties tend to accumulate over time, influencing educational trajectories and transitions.

From a cognitive perspective, bilingual reading comprehension difficulties can be conceptualized as primarily knowledge-based fragilities, whereby higher-level integration processes operate on lexical and morphosyntactic representations that are less rich or less automatized. As a result, word-to-text integration may be slower or less stable, particularly in linguistically dense or abstract

texts (Perfetti and Stafura, 2014). More particularly, when readers encounter constructions that compress information and require reconstruction of argument structure, such as nominalizations, these demands may be further amplified.

Taken together, these two distinct profiles of vulnerability provide a strong rationale for examining event nominalizations in these populations, as such constructions may constitute specific pressure points in reading comprehension.

1.4 Research questions and predictions

The present eye-tracking study aims to address these gaps in the literature by examining two main research questions. First, we examine whether event nominalizations impose greater processing demands than active and passive constructions, and how these demands manifest across accuracy, eye-movement measures, and scanpath patterns. Second, we assess whether such processing costs vary as a function of reader profile, comparing typically developing monolingual children, children with dyslexia, and bilingual children with a migrant background.

Based on the literature reviewed above, we expect nominalizations to increase processing effort relative to active constructions. We also expect them to be more demanding than passive constructions, since both structures involve non-canonical mappings, but nominalizations compress information more strongly and provide fewer explicit cues for argument structure reconstruction. However, such effects may not necessarily translate into reduced comprehension accuracy but rather emerge in online eye-tracking measures. We further predict that children with dyslexia will exhibit elevated overall processing costs, potentially affecting both early and late measures due to reduced decoding automatization and constrained processing resources. In contrast, bilinguals may display comparatively greater integration-related costs, particularly in measures capturing argument reconstruction and global text navigation.

2 Methods

2.1 Participants

Seventy-nine fifth- and sixth-grade children were recruited. Twenty-eight children ($F = 12$) were typically developing monolingual Italian (MON TD: Age $M = 10.85$, SD = 0.69), thirty children ($F = 15$) were typically developing bilingual children acquiring Italian as the societal language (BIL TD: Age $M = 11.01$, SD = 0.78), while twenty-one ($F = 5$) were monolingual Italian children with a diagnosis of DD (MON DD: Age $M = 10.67$, SD = 0.62). Detailed information about the bilingual children's language history and exposure was collected by asking parents to fill in the Q-Bex questionnaire (De Cat et al., 2023), which documented their exposure and use of Italian as societal language and of at least another family language spoken at home. The average age of Italian onset for bilingual children was 3.33 years (SD = 2.82, range = 0–9.11). 80% ($N = 24$) of the bilingual sample was born and raised in Italy, while the others were born in Cuba ($N = 1$), India ($N = 1$), Moldova ($N = 2$), or Pakistan ($N = 2$). Exposure and use estimates were automatically calculated by the Q-BEx online platform from parents' questionnaire responses and are reported here as percentages. Exposure refers to language input received by the child, whereas use refers to the child's own language production within each context. Values are reported for Italian and for each child's most used family language (FL). Since some children were also exposed to and used additional languages, the percentages reported for these two languages do not necessarily sum up to 100%.

The average exposure to Italian at home was 38.74 (SD = 23.81, range = 1.01–100) while use of Italian at home ranged from 2.97 to 100 ($M = 54.55$, SD = 22.37). As for the FL, participants reported highly heterogeneous linguistic backgrounds, including Albanian (1), Arabic (3), English (4), Hindi (1), Portuguese (1), Punjabi (1), Romanian (11), Sinhala (4), Spanish (2), and Urdu (2). Average exposure rate to the FL at home was 53.37 (SD = 21.53, range = 2.14–98.99) while FL use in the home context ranged from 0.99 to 97.03 ($M = 40.56$, SD = 19.60). At school, average exposure to Italian was 89.57 (SD = 19.66, range = 37.24–100) with exposure to the FL being comparatively lower ($M = 7.20$, SD = 16.84, range = 0–62.41). Similarly, participants reported to use Italian ($M = 91.63$, SD = 15.87, range = 35.02–100) comparatively more than their FL in the school context ($M = 5.46$, SD = 10.89, range = 0–39.59). An overview of the exposure to and use of Italian and the FL for each BIL participant is illustrated in Figure 1.

Children with dyslexia were included only if they had a formal diagnosis of developmental dyslexia. Conversely, to minimize the possibility that the two typical groups (both monolinguals and bilinguals) included children with undiagnosed reading difficulties, participants were included only if they had no diagnosis of dyslexia, no teacher-reported marked difficulties in decoding, and did not score below 2 SDs from the age-normed mean in at least two of the following reading measures: word reading accuracy, word reading speed, nonword reading accuracy, and nonword reading speed, which were assessed using the standardized reading tasks DDE-2 (see Materials section). Typically developing children were recruited from two public schools in the area of Verona and Milan (Northern Italy), whereas children with developmental dyslexia were recruited either through the same schools or through local clinical services.

2.2 Materials

2.2.1 Background measures

A series of standardized background tests assessing various linguistic abilities were administered. Word and nonword reading abilities were evaluated using the *DDE-2* (Sartori et al., 2007), considering both reading times and accuracy. To assess receptive vocabulary, participants completed the *Peabody Picture Vocabulary Test–Revised* (PPVT-R; Dunn and Dunn, 2000). Grammatical knowledge was measured through the *Test for Reception of Grammar–Version 2* (TROG-2, Bishop, 2009), whereas reading

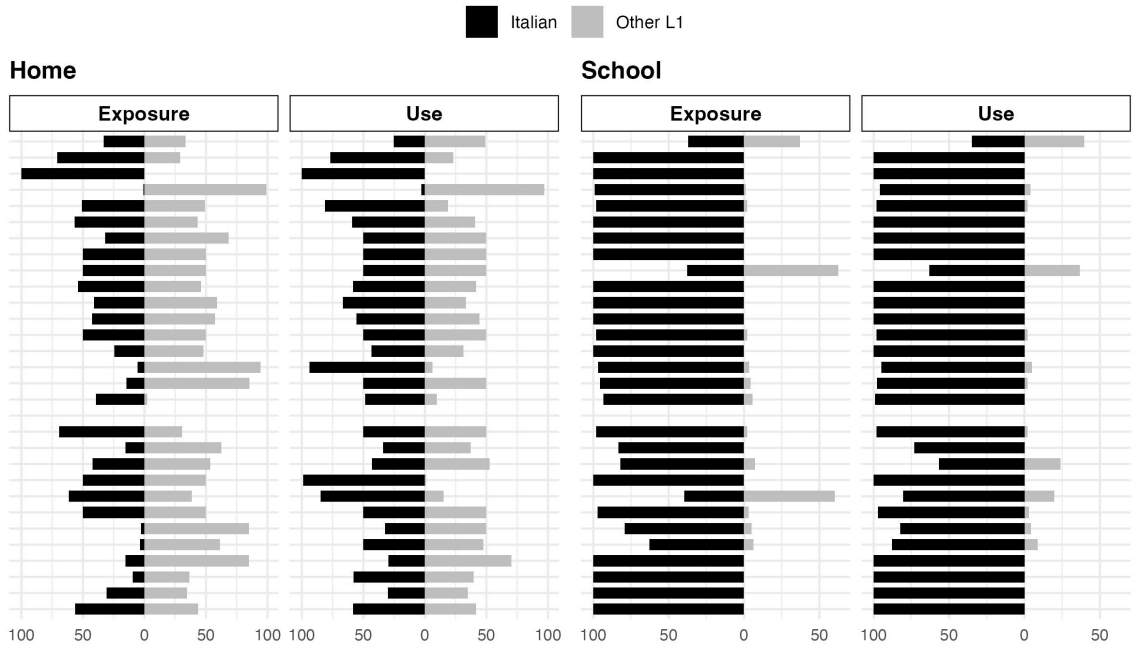


FIGURE 1
Bilingual participants' exposure to and use of Italian and their family language at home (left panel) and at school (right panel).

comprehension was assessed by means of the *Prove MT–Kit scuola battery* (Cornoldi et al., 2017a,b).

2.2.2 Eye-tracking study

A reading comprehension task was developed using Experiment Builder (version 2.5.90, SR Research) to assess participants' processing of texts with nominalized structures and their corresponding active and passive constructions. The task comprised 40 items, four of which served as practice trials. The remaining 36 stimuli consisted of 18 experimental items, six per condition, and 18 fillers. The experimental texts ranged from one to two sentences in length and were drawn from contemporary history, geography, and science school textbooks designed for fifth- and sixth-grade students. The use of textbook-based materials was intended to preserve the relevance of the task to children's reading experiences in real educational settings and to guarantee the ecological validity of our materials. Texts were selected because they contained a transitive verb in a context suitable for adaptation into all three syntactic conditions. Each experimental text was then adapted to create three versions expressing the same event and preserving the critical lexical content as closely as possible, while differing in the syntactic realization of the target event (Nominalization "NOM," Passive construction "PAS," and Active Construction "ACT"). NOM texts featured an event nominalization (e.g., *bombardamento* "bombing"), whereas in PAS or ACT texts, the deverbal noun was replaced with the corresponding verbal structure in passive (e.g., *fu bombardata* "was bombed") or active (e.g., *bombardò* "bombed") voice, respectively (Table 1). Minor adjustments were made to preserve naturalness and semantic equivalence across conditions, ensuring that the critical region did not appear at the beginning or end of the sentence.[1]

To maximize comparability, both the agent and the patient of the target event were explicitly expressed in all versions. However, the three conditions necessarily differed in linear ordering and grammatical realization of the arguments. Specifically, the agent preceded the patient in ACT, whereas in PAS and NOM the patient preceded the agent. In NOM, the patient was expressed as a genitive complement of the event nominal, and the agent was introduced through a *by*-phrase. NOM thus broadly paralleled PAS in argument order and overt agent expression, while differing in the grammatical realization of the patient, which functioned as the subject in PAS.

The nominalizations were not selected to represent a specific morphological type: twelve were derived through the suffix *–(z)ione*, four through conversion, and two through the suffix *–mento*. Lemma frequencies for the nominalized nouns and corresponding base verbs were obtained from the Italian Web 2020 corpus (itTenTen) and are reported in Supplementary Table A2. No significant difference was detected between their frequencies ($\beta = 0.80$, SE $= 0.51$, $t = 1.56$, $p = 0.127$). Fillers were matched to experimental items in length and lexical content but targeted

1 Adjustments concerned how the target construction was integrated into the whole sentence, including minor changes to introductory expressions, the addition of a finite predicate, or modifications to the continuation. For example, referring to the item reported in Table 1, the continuation was expressed with the gerund form *causando* ('causing') in ACT and PAS, whereas in NOM it was expressed with the finite predicate *causò* ('caused') in NOM, with the nominal *il bombardamento* ('the bombing') functioning as its subject. In other cases, where the original passage omitted the agent, it was explicitly introduced so that both event participants were expressed in all versions. See Supplementary Material 1 for a list of all the experimental stimuli.

TABLE 1 Example of one experimental item in all three conditions.

Condition	Example
NOM	Con l'obiettivo di appoggiare le truppe franchiste, nel 1937 **il bombardamento della città basca di Guernica da parte dell'aviazione tedesca** causò centinaia di vittime tra la popolazione civile. *"With the goal of supporting the Francoist forces, in 1937 the bombing of the Basque city of Guernica by the German air force caused hundreds of casualties among the civilian population."*
PAS	Con l'obiettivo di appoggiare le truppe franchiste, nel 1937 **la città basca di Guernica fu bombardata dall'aviazione tedesca**, causando centinaia di vittime tra la popolazione civile. *"With the goal of supporting the Francoist forces, in 1937 the Basque city of Guernica was bombed by the German Air Force, causing hundreds of casualties among the civilian population."*
ACT	Con l'obiettivo di appoggiare le truppe franchiste, nel 1937 **l'aviazione tedesca bombardò la città basca di Guernica**, causando centinaia di vittime tra la popolazione civile. *"With the goal of supporting the Francoist forces, in 1937 the German Air Force bombed the Basque city of Guernica, causing hundreds of casualties among the civilian population."*
Comprehension question	Chi bombardò la città di Guernica? a. L'aviazione tedesca b. Le truppe franchiste c. L'esercito russo
	"Who bombed the city of Guernica? a. *The German Air Force* b. *The Francoist forces* c. *The Russian army"*

The argument structure is highlighted in bold.

different linguistic structures (i.e., anaphoric encapsulation) and were excluded from the analysis.

After reading each text, participants responded to a comprehension question designed to assess their understanding of either the agent or the patient of the target event by selecting one of three response options. Each question included the target answer and two plausible distractors: one mentioned in the text but not corresponding to the requested participant, and one not mentioned in the text. The positions of the target answer and distractors were counterbalanced across the three response options.

The comprehension questions were presented in the active or passive voice in equal proportion. The nine active questions included five targeting the agent, and four targeting the patient, while the nine passive questions included five targeting the patient, and four targeting the agent. Three experimental lists were created by counterbalancing the three syntactic versions (NOM, PAS, and ACT) of the 18 target items, such that each participant saw each experimental text in only one version.

2.3 Procedure

Data collection took place over 2 days in a quiet room at the participating schools or, for most participants with developmental dyslexia, at local health facilities. For the experimental task, participants were seated comfortably in front of a 24-inch monitor connected to an EyeLink 1000 eye-tracker (1,000 Hz) recording monocularly, with their heads stabilized using a chinrest.

Prior to the task, participants were instructed that they would read a series of short texts on history, geography, or science. They were asked to read each text silently before clicking the mouse to display a comprehension question. Although the text remained available after clicking the mouse (Screen 2), participants were reminded to read it carefully during its initial presentation (Screen 1). The texts were presented in a monospaced font (Courier New, size 24, black); for drift-correction purposes, each trial was preceded by a fixation cross. The mouse click marked the end of Screen 1 and the beginning of Screen 2, where the comprehension question and response options appeared; the click was not included in the reading-time measure. Participants could read the text for as long as needed, as no maximum viewing duration was imposed.

The experimental task was administered across two sessions, each including 18 stimuli (nine experimental and nine fillers) and an optional halfway break. The second session took place approximately 2 weeks after the first. Experimental items were randomly assigned to the two sessions, with the constraint that each session included three items per condition and nine fillers. Presentation order and syntactic structure type were also counterbalanced across lists.

Each eye-tracking session lasted about 15 min and began with a standard 9-point calibration and validation procedure, followed by practice trials with automated feedback. Eye movements were recorded monocularly, using the right eye by default. In a small number of cases, the left eye was selected when it provided better calibration quality.

In both sessions, the experimental task was followed by the administration of the background tasks described above (DDE, PPVT-R, TROG, MT comprehension tasks), bringing the total

duration of each session to approximately 45 min. For bilingual participants, parents completed questionnaires regarding their children's language use and exposure.

The study received ethical approval from the Research Ethics Committee at the University of Verona. Written informed consent was obtained from the children's parents, and children provided assent prior to participation.

3 Results

3.1 Data cleaning

During recording, blink and loss-of-tracking events were automatically flagged by the eye-tracking software; no additional manual blink- or track-loss threshold was applied during preprocessing. A drift-correction fixation cross preceded each trial, and recalibration was performed whenever necessary. Eye-tracking data cleaning comprised four steps. First, fixations shorter than 80 ms were merged with adjacent fixations if the distance between the two was smaller than 0.5 degrees. Second, all fixations shorter than 80 ms or longer than 1,000 ms were removed. Third, all fixations falling outside the AOI relevant for each analysis were also removed. Finally, a visual inspection was carried out by two authors independently and then compared. This was done to remove poorly calibrated trials, and those in which Screen 1 was not entirely read. This led to the removal of 728 trials (26%), and two bilinguals. The analyzed sample thus included twenty-eight TD bilinguals, twenty-eight TD monolinguals, and twenty-one DD monolinguals.

3.2 Data analysis

The data analysis comprised three steps: first, accuracy scores were analyzed through a generalized mixed effect model (*glmer* function, *lme4* package, Bates et al., 2015) in R (version 4.5.3). Second, we ran a *scanpath analysis* focusing specifically on the argument structure (AS) for Screen 1, and on the text, question and response options for Screen 2; for the latter, we focused on correctly answered trials to characterize the reading patterns leading to successful retrieval of information about event participants. Given that similar reading patterns may accompany different response outcomes, this restriction allowed us to characterize how children navigated the text when they correctly identified the requested event participant.

The scanpath analyses proceeded as follows: a *dissimilarity* score was calculated for each item separately through the *scasim* function (*scanpath* package, von der Malsburg, 2018). We used the dissimilarity score as an index of *scanpath regularity,* as it measures how regular the reading strategy for a given trial is as compared to the most regular reading strategy (i.e., the central scanpath). Then, we calculated scanpath *density* through the *isoMSD* function (*MASS* package, Venables and Ripley, 2002) and ran a cluster analysis through the *densityMclust* function (*mclust* package, Scrucca et al., 2023). The cluster analysis–performed on Screen 2 only–was used to identify the different reading strategies employed by participants to retrieve the information required to answer the comprehension question.

The final step concerned the analysis of *total dwell time* and the total number of *regressions in*. For both measures, the target AOI comprised the text displayed on each screen, including the question and response options on Screen 2. *Total dwell time* was calculated by summing the dwell times across all words in the text, whereas the total number of *regressions in* was calculated by summing the number of regressions into each word. Following the definition used in Data Viewer (SR Research), a *regression in* refers to a regressive eye movement into a given interest area from a later interest area. Thus, the resulting measure captured the total number of backward eye movements into previously encountered portions of the text. The full list of experimental items and the areas of interest used in the analyses are provided in Supplementary material.

3.3 Background measures

Participants' performance in the background measures is reported in Table 2. Typically developing children–both bilinguals and monolinguals–outperformed children with dyslexia in all reading measures. By contrast, differences between TD monolinguals and bilinguals were not significant in terms of lexical and sublexical reading, but TD monolingual children outperformed TD bilingual children in reading comprehension. Further, TD monolinguals outperformed both TD bilinguals and children with dyslexia in the PPVT-R task tapping receptive vocabulary; no differences emerged between TD bilinguals and children with DD in this case. No difference across groups emerged in the TROG task tapping grammatical knowledge.

3.4 Eye-tracking study-accuracy

Participants' responses were binary-coded (1: Correct, 0: Incorrect) and analyzed as a function of a two-way interaction between Condition and Group, controlling for Question Type (ACT-Agent, ACT-Patient, PAS-Agent, PAS-Patient) as well as for participants' performance in the TROG and PPVT-R tests. Random intercepts were specified for Participants and Items, with a random slope for Condition included for both Participants and Items. The model was fitted using a binomial family, with *bobyqa* as the optimizer. Following removal of the by-participant random slope for Condition due to singularity, the resulting model was evaluated for the contribution of the Group × Condition interaction, as well as of Question Type, TROG, and PPVT-R using likelihood-ratio tests (*drop1* function, *stats* package, R Core Team, 2025). Because this interaction did not provide evidence of improved model fit, it was removed, and the more parsimonious model was retained. Question Type was also dropped as it did not improve model fit. Simulation-based residual diagnostics (*DHARMa* package; Hartig, 2026) indicated no evidence of overdispersion (dispersion = 1.01, $p = 0.928$).

The best-fit model yielded a significant main effect of Group ($\chi^2(2) = 13.77$, $p = 0.001$), Condition ($\chi^2(2) = 8.03$, $p = 0.018$),

TABLE 2 Background tests: mean and SD are reported, along with *emmeans* contrasts derived from simple linear models (formula = *DV* ~ Group).

Descriptive statistics			Inferential statistics				
TD	DD	BIL	Contrasts	β	SE	t	p
Word reading time							
0.33 (0.66)	−2.57 (2.46)	−0.15 (0.88)	BIL_TD - MON_DD	2.42	0.41	5.87	<0.001
			BIL_TD - MON_TD	−0.48	0.38	−1.27	0.626
			MON_DD - MON_TD	−2.90	0.41	−6.99	<0.001
Word reading errors							
0.43 (0.48)	−2.21 (1.43)	0.13 (0.87)	BIL_TD - MON_DD	2.33	0.27	8.61	<0.001
			BIL_TD - MON_TD	−0.30	0.25	−1.20	0.456
			MON_DD - MON_TD	−2.63	0.27	−9.58	<0.001
Pseudoword reading time							
0.42 (0.75)	−1.38 (1.45)	0.16 (0.68)	BIL_TD - MON_DD	1.53	0.28	5.51	<0.001
			BIL_TD - MON_TD	−0.26	0.25	−1.02	0.569
			MON_DD - MON_TD	−1.79	0.28	−6.39	<0.001
Pseudoword reading errors							
0.23 (0.77)	−0.88 (1.32)	0.15 (0.85)	BIL_TD - MON_DD	1.03	0.28	3.68	0.001
			BIL_TD - MON_TD	−0.08	0.25	−0.30	0.952
			MON_DD - MON_TD	−1.11	0.28	−3.90	0.001
Reading comprehension (MT)							
0.46 (0.77)	−1.75 (0.84)	−0.33 (1.08)	BIL_TD - MON_DD	1.42	0.28	5.07	<0.001
			BIL_TD - MON_TD	−0.79	0.25	−3.21	0.006
			MON_DD - MON_TD	−2.21	0.29	−7.72	<0.001
Receptive Vocabulary (PPVT-R)							
104.75 (12.69)	92.2 (10.59)	91.9 (16.69)	BIL_TD - MON_DD	−0.30	4.02	−0.07	0.997
			BIL_TD - MON_TD	−12.85	3.66	−3.51	0.002
			MON_DD - MON_TD	−12.55	4.08	−3.08	0.008
Grammatical knowledge (TROG)							
100.96 (16)	91.45 (13.52)	93.6 (11.07)	BIL_TD - MON_DD	2.15	3.94	0.55	0.849
			BIL_TD - MON_TD	−7.36	3.58	−2.06	0.106
			MON_DD - MON_TD	−9.51	3.99	−2.38	0.051

For reading measures, standardized scores are shown, where 0 corresponds to the average score based on normative data. For all standardized reading measures, including reading time and error scores, higher values indicate better performance. Positive values indicate above-average performance, whereas negative values indicate below-average performance relative to the normative sample.

TROG ($\chi^2(1) = 4.50$, $p = 0.034$), and PPVT-R ($\chi^2(1) = 7.30$, $p = 0.007$), indicating that participants with higher grammatical and vocabulary skills were overall more accurate. *Post-hoc* analyses were run on the Group and Condition effects via the *emmeans* function (Lenth and Piaskowski, 2025) with Bonferroni adjustment for multiple comparisons. BIL TD were significantly more accurate than MON DD ($\beta = 0.78$, SE = 0.21, 95% CI [0.27, 1.28], $z = 3.70$, $p = 0.001$), but no differences emerged between MON TD and BIL TD ($\beta = 0.32$, SE = 0.22, 95% CI [−0.21, 0.85], $z = 1.44$, $p = 0.447$), or between MON TD and MON DD ($\beta = -0.46$, SE = 0.22, 95% CI [−0.99, 0.08], $z = -2.05$, $p = 0.122$). The *post-hoc* analysis on the Condition effect yielded significant differences between ACT and PAS ($\beta = -0.63$, SE = 0.23, 95% CI [−1.17, −0.09], $z = -2.77$, $p = 0.017$), indicating that ACT texts were more difficult to respond to compared to PAS texts; no difference emerged between ACT and NOM ($\beta = -0.18$, SE = 0.28, 95% CI [−0.85, 0.49], $z = -0.65$, $p = 1.000$) or between PAS and NOM ($\beta = -0.45$, SE = 0.28, 95% CI [−1.11, 0.21], $z = -1.63$, $p = 0.312$). The model output is plotted in Figure 2.

3.5 Scanpath analysis

3.5.1 Screen 1

The *dissimilarity* score was modeled via *lmer* as a function of a two-way interaction between Condition and Group, controlling for PPVT-R and TROG scores. In this case, Question Type was

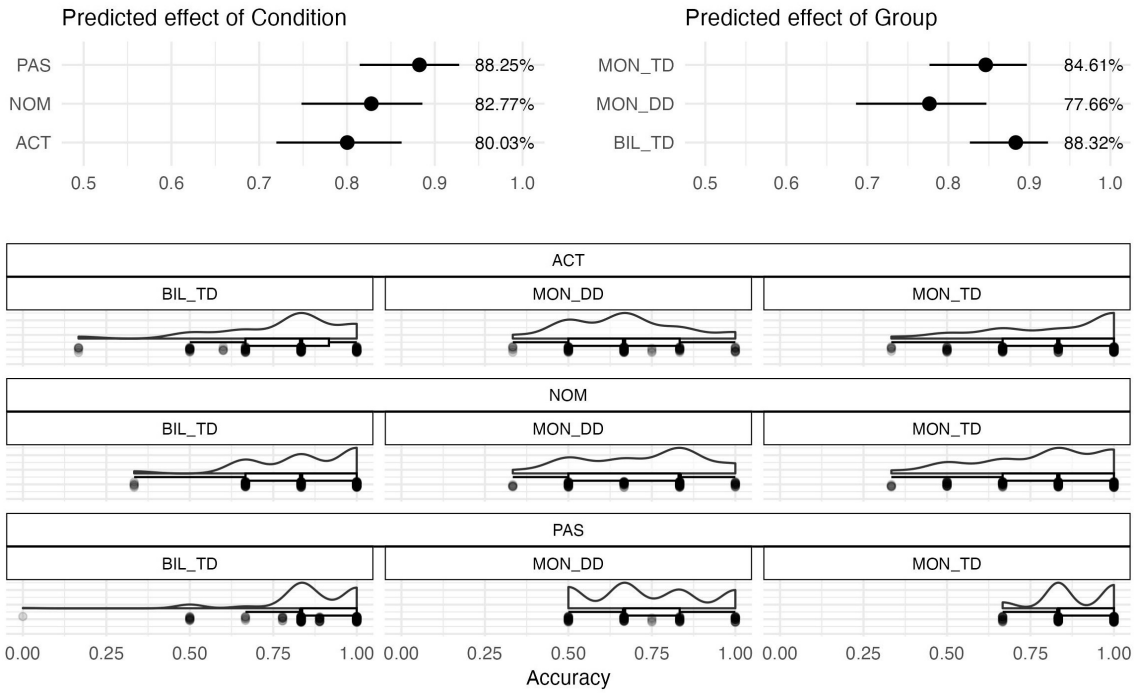


FIGURE 2
Main effects of Condition and Group for accuracy scores (upper panel): estimated marginal means are reported for each Group and Condition. In the lower panel, points represent individual observed accuracy values, while the density curves show the distribution of accuracy scores within each Condition and Group.

not included in the model because on *Screen 1* participants had not been presented with the comprehension question yet. Random intercepts were specified for Participants and Items, with a random slope for Condition included for both intercepts. Random slopes associated with singular fits were removed. The two-way interaction was also excluded, as the more parsimonious model provided a better fit to the data, as indicated by its lower AIC (interaction model: AIC = 1,771.1; reduced model: AIC = 1,766.1). TROG and PPVT-R were also dropped as they did not contribute to better model fit. The best-fit model was specified as follows: *dissimilarity* ~ *condition* + *group* + (1| *participant*) + (1|*item*). Simulation-based residual diagnostics indicated no evidence of overdispersion (dispersion = 1.03, $p = 0.344$).

The model yielded significant effects of Condition ($\chi^2(2) = 55.32$, $p < 0.001$) and Group ($\chi^2(2) = 71.30$, $p < 0.001$). *Post-hoc* analyses were run on the Group and Condition effects via the *emmeans* function (Lenth and Piaskowski, 2025) with Bonferroni adjustment for multiple comparisons. Dissimilarity scores were significantly higher for MON DD compared to both MON TD ($\beta = 1.17$, SE = 0.15, 95% CI [0.81, 1.53], $t = 7.91$, $p < 0.001$) and BIL TD ($\beta = -1.04$, SE = 0.15, 95% CI [−1.40, −0.67], $t = -6.99$, $p < 0.001$); the difference between MON TD and BIL TD was not significant ($\beta = 0.13$, SE = 0.14, 95% CI [−0.20, 0.47], $t = 0.98$, $p = 0.992$). As for Condition, the *post-hoc* analyses yielded significantly higher dissimilarity scores for NOM as compared to both ACT ($\beta = -0.30$, SE = 0.04, 95% CI [−0.40, −0.19], $t = -6.86$, $p < 0.001$) and PAS ($\beta = 0.25$, SE = 0.04, 95% CI [0.15, 0.35], $t = 5.86$, $p < 0.001$); the difference between ACT and PAS was not significant ($\beta = -0.05$, SE = 0.04, 95% CI [−0.15, 0.06], $t = -1.12$, $p = 0.790$). The model output is illustrated in Figure 3.

3.5.2 Screen 2

Dissimilarity scores for Screen 2 (AOI: text + question, accurate trials only) were modeled as a function of the Group × Condition interaction, controlling for Question Type, PPVT-R, and TROG. The initial model included the maximal random-effects structure with random intercepts for Participants and Items and by-participant and by-item random slopes for Condition. Due to singularity, the random-effects structure was simplified by removing both random slopes and the by-item random intercept, whose variance was estimated at zero. The interaction between Group and Condition was also dropped as it did not improve model fit (interaction model: AIC = 1,888.5; reduced model: AIC = 1,885.0). The fixed effect of TROG, and PPVT-R, and Question Type were also dropped. Simulation-based residual diagnostics indicated no evidence of overdispersion (dispersion = 1.05, $p = 0.328$). The best-fit model yielded a significant main effect of Condition ($\chi^2(2) = 7.40$, $p = 0.025$) and Group ($\chi^2(2) = 11.95$, $p = 0.003$). *Post-hoc* analyses (*emmeans* function, Bonferroni adjustment) on the Condition effect indicated that dissimilarity scores were significantly higher for NOM as compared to PAS ($\beta = 0.17$, SE = 0.06, 95% CI [0.01, 0.32], $t = 2.63$, $p = 0.026$); no significant differences emerged between ACT and PAS ($\beta = 0.04$, SE = 0.06, 95% CI [−0.11, 0.20], $t = 0.63$, $p = 1.000$) or ACT and NOM ($\beta = -0.13$, SE = 0.07, 95% CI [−0.28, 0.03],

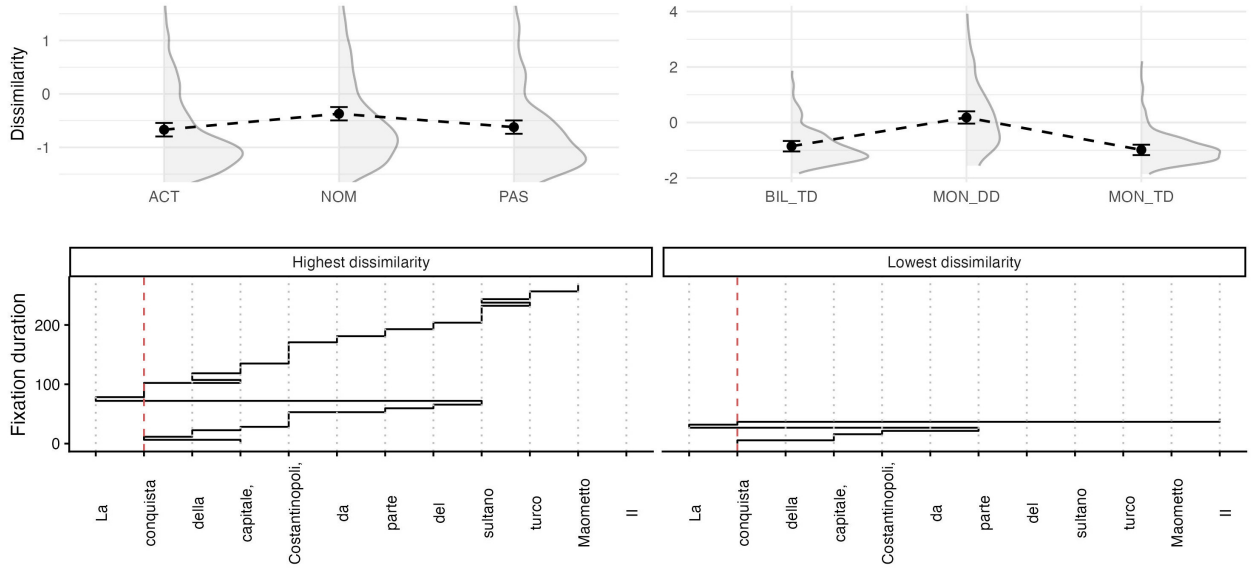


FIGURE 3
Main effects of Condition and Group on the dissimilarity score (upper panel); example of high (bottom left panel) and low dissimilarity trials (bottom right panel). AOI: Screen 1. In the bottom panel, the y axis illustrates the total fixation duration (ms) on the trial.

$t = -1.91$, $p = 0.168$). *Post-hoc* comparisons on the Group effect yielded significant differences between MON TD and MON DD ($\beta = 0.53$, SE = 0.16, 95% CI [0.15, 0.92], $t = 3.41$, $p = 0.003$); no difference emerged between MON TD and BIL TD ($\beta = 0.29$, SE = 0.14, 95% CI [−0.06, 0.64], $t = 2.05$, $p = 0.131$) or between BIL TD and MON DD ($\beta = -0.24$, SE = 0.16, 95% CI [−0.63, 0.14], $t = -1.53$, $p = 0.387$), as illustrated in Figure 4.

Finally, a *cluster analysis* (AOI: text + question, accurate trials only) was performed to identify the reading strategies used by participants to retrieve the information relevant to answer the comprehension question. Given the inherently subjective nature of assigning clusters to specific reading strategies, this classification was performed independently by four authors. The resulting assignments were then compared, and inter-rater reliability (IRR) was calculated to assess the consistency of the classifications using the *kappam.fleiss* function from the *irr* package in R (Gamer et al., 2019). Three different reading strategies were identified: a (i) *minimal rescanning,* with reading concentrated mainly on the question and response options and little return to the text; (ii) *partial rescanning,* involving a more sustained return to the text, particularly in the region containing the event and its arguments, with limited regressions and re-readings; and a (iii) *maximal rescanning*, involving prolonged re-reading and repeated back-and-forth movements across the text, as illustrated in Figure 5 (panel B). The IRR for the three strategies was good (minimal rescanning: $k = 0.87$, $z = 11.1$, $p < 0.001$; partial rescanning: $k = 0.83$, $z = 12.00$, $p < 0.001$; maximal rescanning: $k = 0.89$, $z = 10.4$, $p < 0.001$).

The probability for each strategy was tested as a function of Condition and Group via the *mblogit* function (*mclogit* package, Elff, 2025). The random effect structure was kept maximal and then compared with a model including only by-item and by-participant intercepts. The simpler model had a better AIC (1,500.6 vs. 1,575.3 for model with random slopes for Condition) and was therefore used to estimate *post-hoc* comparisons (*emmeans* function, Bonferroni adjustment) for the Condition and Group effects. Notably, the probability of a *minimal rescanning* strategy was higher for ACT as compared to NOM ($\beta = 0.12$, SE = 0.05, 95% CI [0.00, 0.23], $z = 2.41$, $p = 0.048$) and, by contrast, the probability of a *partial rescanning* was higher for NOM as compared to ACT ($\beta = -0.12$, SE = 0.05, 95% CI [−0.23, −0.01], $z = -2.55$, $p = 0.033$). *Post-hoc* comparisons for the Condition effect are reported in Table 3.

As for Group, *post-hoc* comparisons suggest a higher probability of *maximal rescanning* for MON DD compared to MON TD ($\beta = 0.19$, SE = 0.06, 95% CI [0.05, 0.33], $z = 3.25$, $p = 0.003$); no difference emerged between BIL TD and MON DD ($\beta = -0.14$, SE = 0.06, 95% CI [−0.28, 0.01], $z = -2.20$, $p = 0.071$) or between MON TD and BIL TD ($\beta = 0.05$, SE = 0.04, 95% CI [−0.03, 0.14], $z = 1.53$, $p = 0.278$). By contrast, the probability of a *minimal rescanning* was higher for MON TD compared to both MON DD ($\beta = -0.29$, SE = 0.09, 95% CI [−0.51, −0.07], $z = -3.15$, $p = 0.005$) and BIL TD ($\beta = -0.21$, SE = 0.09, 95% CI [−0.41, −0.01], $z = -2.48$, $p = 0.035$); no difference emerged between MON DD and BIL TD ($\beta = 0.08$, SE = 0.09, 95% CI [−0.14, 0.3], $z = 0.84$, $p = 0.667$). *Post-hoc* comparisons on the Group effect are reported in Table 4 (see also Figure 5, panel A).

3.6 Aggregate measures

3.6.1 Screen 1

Log-transformed total dwell time was used as a proxy for reading time and analyzed using a *lmer* model with a two-way interaction between Condition and Group and controlling for TROG, PPVT-R, and word count (AOI Length). The random effect structure included random intercepts for Participants and Items, and by-participant and by-item slopes for Condition. Due

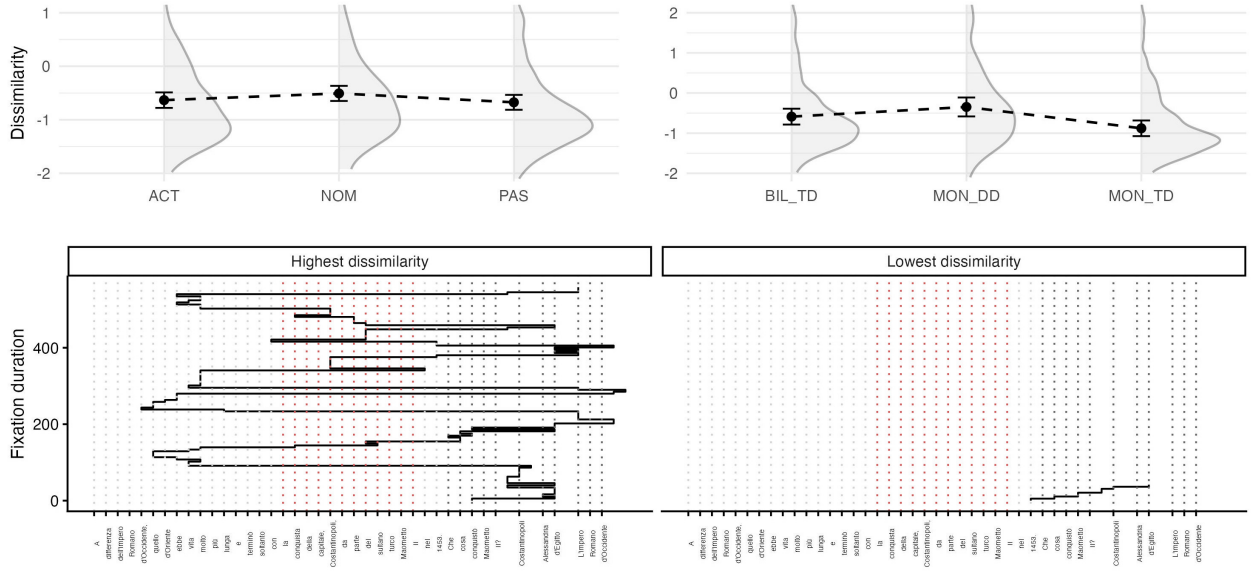


FIGURE 4
Main effect of Condition and Group on the dissimilarity score in the scanpath analysis (upper panel). AOI: Screen 2 (Text + Question; correct trials only). The lower panels illustrate scanpaths with high **(left)** and low **(right)** dissimilarity. The high-dissimilarity example shows repeated returns to previously read portions of the text, whereas the low-dissimilarity example shows a brief sequence of fixations on the question and the response options.

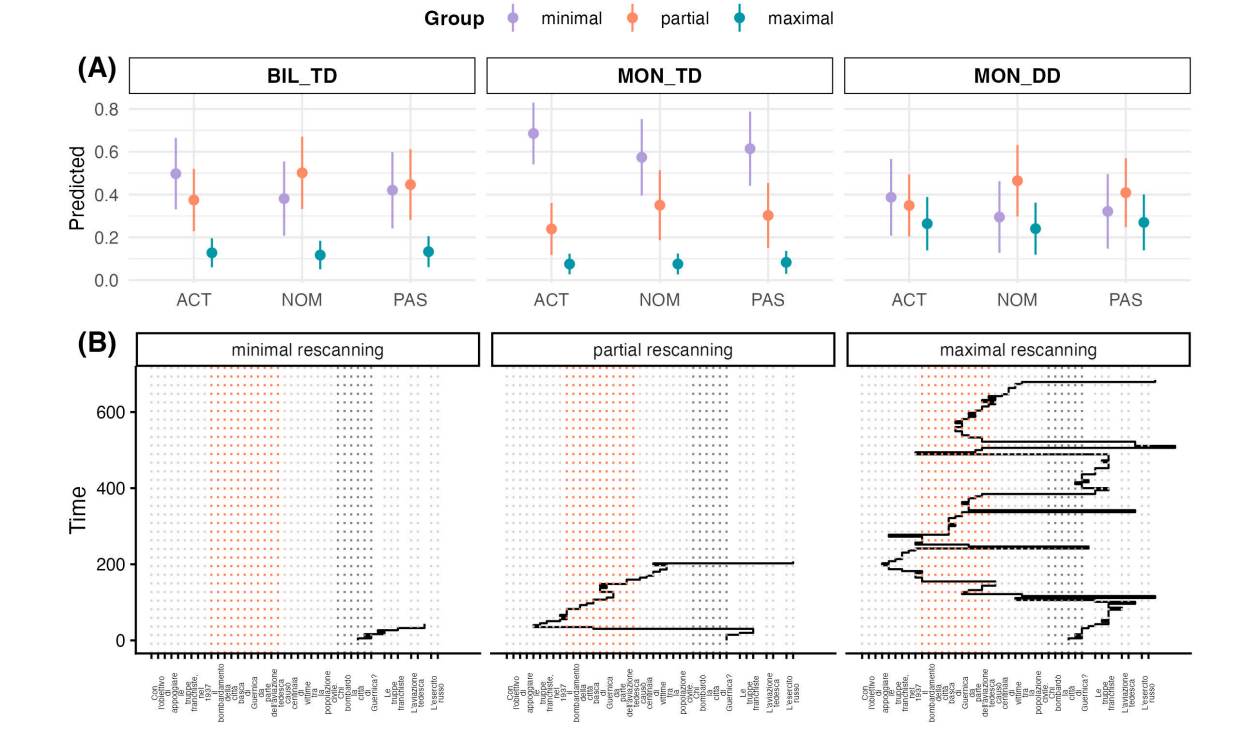


FIGURE 5
Probability of maximal, partial, or minimal rescanning by Condition and Group **(A)**; reading strategies are illustrated in **(B)**. In panel B, orange dashed lines identify the words belonging to the target event's argument structure.

to singularity, both slopes were dropped. After model selection was performed, the best-fit model was specified as follows: *total dwell* $\sim$ *condition* + *group* + *AOI length* + (1 | *Participant*) + (1 | *Item*). Simulation-based residual diagnostics indicated no evidence of overdispersion (dispersion = 1.03, $p = 0.504$). The model yielded a significant main effect of Group ($\chi^2(2) = 66.74$, $p < 0.001$), indicating longer dwell time for MON DD compared to both BIL TD ($\beta = -0.60$, SE = 0.09, 95% CI [−0.83, −0.38],

TABLE 3 *Post-hoc* comparisons on the probability of each reading strategy by condition.

Contrasts	β	SE	95% CI	z	p
ACT					
Maximal - Minimal	−0.38	0.08	[−0.58, −0.18]	−4.65	<0.001
Maximal - Partial	−0.17	0.06	[−0.32, −0.02]	−2.78	0.016
Minimal - Partial	0.21	0.11	[−0.06, 0.48]	1.85	0.195
NOM					
Maximal - Minimal	−0.26	0.07	[−0.44, −0.09]	−3.55	0.001
Maximal - Partial	−0.29	0.07	[−0.45, −0.12]	−4.15	<0.001
Minimal - Partial	−0.02	0.12	[−0.3, 0.25]	−0.20	1.000
PAS					
Maximal - Minimal	−0.29	0.08	[−0.47, −0.1]	−3.68	0.001
Maximal - Partial	−0.22	0.07	[−0.38, −0.06]	−3.26	0.003
Minimal - Partial	0.07	0.11	[−0.2, 0.34]	0.62	1.000
Minimal rescanning					
ACT - NOM	0.12	0.05	[0, 0.23]	2.41	0.048
ACT - PAS	0.08	0.05	[−0.04, 0.19]	1.65	0.297
NOM - PAS	−0.04	0.05	[−0.15, 0.07]	−0.84	1.000
Partial rescanning					
ACT - NOM	−0.12	0.05	[−0.23, −0.01]	−2.55	0.033
ACT - PAS	−0.06	0.04	[−0.17, 0.04]	−1.41	0.473
NOM - PAS	0.05	0.05	[−0.05, 0.16]	1.21	0.680
Maximal rescanning					
ACT - NOM	0.00	0.03	[−0.08, 0.08]	−0.02	1.000
ACT - PAS	−0.02	0.03	[−0.1, 0.06]	−0.51	1.000
NOM - PAS	−0.02	0.03	[−0.09, 0.06]	−0.49	1.000

$t = -6.55$, $p < 0.001$) and MON TD ($\beta = 0.72$, SE = 0.09, 95% CI [0.49, 0.94], $t = 7.77$, $p < 0.001$); no difference emerged between MON TD and BIL TD ($\beta = 0.11$, SE = 0.08, 95% CI [−0.10, 0.32], $t = 1.31$, $p = 0.579$). The effect of AOI Length was also significant ($\chi^2(1) = 61.35$, $p < 0.001$); Condition was not significant ($\chi^2(2) = 0.79$, $p = 0.673$).

The number of *regressions "in"* was modeled with the *glmer.nb* function; the initial model structure was the same specified above. Due to convergence issues, by-item and by-participant random slopes for Condition were dropped (best-fit model dispersion = 1.10, $p = 0.192$). The likelihood ratio tested indicated that the interaction between Group and Condition (LRT = 2.17, $p = 0.705$), as well as TROG (LRT = 0.02, $p = 0.874$) and PPVT-R (LRT = 0.43, $p = 0.5113$) did not improve model fit. The best-fit model thus only included main effects of Condition ($\chi^2(2) = 3.83$, $p = 0.147$), Group ($\chi^2(2) = 7.40$, $p = 0.025$), and Word Count ($\chi^2(1) = 52.65$, $p < 0.001$). *Post-hoc* comparisons on the Group effect indicated that MON DD made a higher number of regressions compared to MON TD ($\beta = 0.31$, SE = 0.11, 95% CI [0.04, 0.59], z = 2.72, $p = 0.020$) but no difference between MON DD and BIL TD ($\beta = -0.19$, SE = 0.11, 95% CI [−0.46, 0.09], z = −1.63, $p = 0.312$) or between MON TD and BIL TD ($\beta = 0.12$, SE = 0.11, 95% CI [−0.13, 0.38], z = 1.18, $p = 0.712$).

3.6.2 Screen 2

Log-transformed total dwell time was analyzed using a *lmer* model with a two-way interaction between Condition and Group and controlling for Question Type, TROG, PPVT-R, and word count (AOI Length). The random effect structure included random intercepts for Participants and Items, and by-participant and by-item slopes for Condition. Due to singularity, the by-item slope was dropped. After model selection was performed, the best-fit model was specified as follows: *total dwell* ∼ *condition* + *group* + *question type* + *AOI length* + (1 + *Condition*| *Participant*) + (1 | *Item*). Simulation-based residual diagnostics indicated no evidence of overdispersion (dispersion = 1.06, $p = 0.128$). The model yielded a significant main effect of Group ($\chi^2(2) = 17.88$, $p < 0.001$); Condition ($\chi^2(2) = 4.31$, $p = 0.116$), Question Type ($\chi^2(3) = 7.45$, $p = 0.059$) and AOI Length ($\chi^2(1) = 0.50$, $p = 0.480$) were not significant. *Post-hoc* analyses on the Group effect indicated significantly longer dwell time for both MON DD ($\beta = 0.53$, SE = 0.13, 95% CI [0.21, 0.85], $t = 4.05$, $p < 0.001$) and BIL TD ($\beta = 0.29$, SE = 0.13, 95% CI [0.00, 0.59], $t = 2.44$, $p = 0.051$) compared to MON TD, although the latter was only marginal. No difference emerged between MON DD and BIL TD ($\beta = -0.24$, SE = 0.13, 95% CI [−0.56, 0.08], z = −1.80, $p = 0.227$).

TABLE 4 *Post-hoc* comparisons on the probability of each reading strategy by group.

Contrasts	β	SE	95% CI	z	p
Minimal rescanning					
BIL TD - MON DD	0.08	0.09	[−0.14, 0.3]	0.84	0.677
BIL TD - MON TD	−0.21	0.09	[−0.41, −0.01]	−2.48	0.035
MON DD - MON TD	−0.29	0.09	[−0.51, −0.07]	−3.15	0.005
Partial rescanning					
BIL TD - MON DD	0.06	0.08	[−0.13, 0.24]	0.71	0.756
BIL TD - MON TD	0.16	0.07	[−0.01, 0.33]	2.15	0.081
MON DD - MON TD	0.10	0.08	[−0.08, 0.28]	1.32	0.382
Maximal rescanning					
BIL TD - MON DD	−0.14	0.06	[−0.28, 0.01]	−2.20	0.071
BIL TD - MON TD	0.05	0.04	[−0.03, 0.14]	1.53	0.278
MON DD - MON TD	0.19	0.06	[0.05, 0.33]	3.25	0.003
BIL TD					
maximal - minimal	−0.29	0.09	[−0.51, −0.08]	−3.20	0.004
maximal - partial	−0.32	0.07	[−0.5, −0.15]	−4.37	<0.001
minimal - partial	−0.03	0.14	[−0.36,0.3]	−0.21	0.975
MON DD					
maximal - minimal	−0.08	0.12	[−0.36, 0.2]	−0.66	0.786
maximal - partial	−0.13	0.10	[−0.36, 0.09]	−1.38	0.349
minimal - partial	−0.05	0.14	[−0.38, 0.27]	−0.39	0.921
MON TD					
maximal - minimal	−0.56	0.09	[−0.76, −0.36]	−6.55	<0.001
maximal - partial	−0.22	0.06	[−0.36, −0.08]	−3.71	0.001
minimal - partial	0.34	0.13	[0.03, 0.65]	2.54	0.030

Following the approach described above, the total number of *regressions* was modeled as a function of a two-way interaction between Condition and Group, controlling for Question Type, PPVT-R, TROG, and AOI Length. The random effect structure included random intercepts for Participants and Items, allowing for by-participant and by-item slopes for Condition. Due to convergence issues, random slopes were dropped. Single-term deletion tests indicated that the interaction between Condition and Group (LRT = 4.01, $p = 0.404$) did not improve the model fit and its removal resulted in lower AIC (5,934 vs. 5,938). TROG (LRT = 1.64, $p = 0.200$) and PPVT-R (LRT = 1.13, $p = 0.286$) were also dropped as they did not improve model fit. Simulation-based residual diagnostics indicated no evidence of overdispersion (dispersion = 1.16, $p = 0.096$). The best-fit model yielded a significant main effect of Group ($\chi^2(2) = 6.55, p = 0.038$), Question Type ($\chi^2(3) = 11.22, p = 0.011$), and AOI Length ($\chi^2(1) = 6.86, p = 0.009$). Condition was not significant ($\chi^2(2) = 4.35, p = 0.114$). *Post-hoc* analyses (with Bonferroni correction) revealed that BIL TD made a significantly higher number of regressions than MON TD ($\beta = 0.33$, SE = 0.13, 95% CI [0.02, 0.64], z = 2.54, $p = 0.033$); no other group difference emerged (MON TD vs. MON DD: $\beta = 0.21$, SE = 0.14, 95% CI [−0.14, 0.55], z = 1.44, $p = 0.448$; MON DD vs. BIL TD: $\beta = 0.13$, SE = 0.14, 95% CI [−0.22, 0.47], z = 0.87, $p = 1.000$). As for Question Type, *post-hoc* analyses indicated that active questions targeting the agent received more regressions compared to both active questions targeting the patient ($\beta = 0.35$, SE = 0.12, 95% CI [0.04, 0.67], z = 2.95, $p = 0.019$) and passive questions targeting the patient ($\beta = 0.31$, SE = 0.11, 95% CI [0.02, 0.61], z = 2.79, $p = 0.032$). No other significant contrast emerged.

4 Discussion

The present study investigated how event nominalizations affect reading processing and comprehension in developing readers, focusing specifically on three reader profiles: typically developing monolingual children, children with dyslexia, and bilingual children with Italian as the societal language. Using eye-tracking measures, we examined how texts containing nominalizations are processed with respect to semantically equivalent active and passive constructions and whether these effects vary depending on the reader profile. Overall, our findings

indicate that nominalizations affect online processing, as reflected in scanpath dissimilarity and rereading strategies, with distinct patterns emerging across reader groups.

4.1 Nominalizations and processing demand in developing readers

Our first research question concerned whether nominalizations impose greater processing demands than semantically equivalent active and passive constructions. Our results show that they do not reduce comprehension accuracy compared to other conditions. However, eye-tracking measures, and particularly the scanpath analyses, revealed that the use of nominals triggers differences in reading patterns during both initial reading and question-answering. This pattern is consistent with previous studies in adult populations, showing that nominalizations increase processing cost without affecting comprehension accuracy (Hansen et al., 2006; Wolfer, 2016). Crucially, our results extend these findings to developing readers and indicate that the cognitive cost associated with nominalizations mainly emerges in online processing measures rather than in accuracy. Within this scenario, the scanpath analyses provided particularly interesting information on how readers navigated the text, offering a window into the regularity and organization of their eye movements. More specifically, during initial reading (Screen 1), we found higher dissimilarity in texts containing nominals compared to both verbal constructions, which suggests that readers are less consistent in how they distribute attention across the text. This can be interpreted as evidence of increased processing difficulty or uncertainty (Mézière et al., 2024), thus pointing to greater demands in the first processing of the nominalized event structure, even before the comprehension question appeared. During question-answering (Screen 2), analysis of correctly answered trials reported higher scanpath dissimilarity for nominalizations than for passives, indicating higher efforts accompanying the identification of the appropriate answer. The cluster analysis was particularly informative: compared to active texts, in fact, nominalized texts were associated with a higher probability of partial rescanning and a lower probability of minimal rescanning. Importantly, partial rescanning involved a more sustained return to specific portions of the text, including the region containing the event and its arguments, whereas minimal rescanning involved little return to the text beyond reading the question and response options. Considering that all comprehension questions targeted event participants, this pattern is consistent with greater effort triggered by nominalized texts in reconstructing the event and retrieving information about its participants compared to active texts.

We interpret this pattern in light of the structural properties of nominalizations: as discussed in the Introduction, nominalizations compress event structure into a nominal phrase, reducing the transparency of thematic role assignment and requiring readers to reconstruct predicate-argument relations and retrieve implicit information from context (Bello, 2016; Ravelli, 1996; Schleppegrell, 2004). The relevant processing demand triggered by nominals therefore concerns the whole organization of the event information within the sentence and its integration into the unfolding text representation. Although both the agent and the patient were explicitly expressed in our materials, their thematic roles were not mapped onto canonical syntactic positions in nominalized constructions, requiring readers to reconstruct the predicate–argument relations associated with the nominalized event and integrate its participants into the text representation. The reading patterns associated with nominalized texts are thus consistent with additional demands in recovering these relations and constructing a representation of the event.

The comparison with both verbal constructions helps clarify this interpretation. From the structural point of view, it is worth noting that active sentences preserve canonical mappings between syntactic position and thematic roles, while passive constructions involve non-canonical mappings but retain explicit morphosyntactic cues that may support comprehension. During initial reading, both elicited lower scanpath dissimilarity than nominalizations, without significant differences between them. The contrast between passives and nominalizations is particularly informative since they share the non-canonical word order and the overt expression of the agent through a by-phrase. The fact that they differ in dissimilarity, also observed during question-answering, suggests that the demands imposed by nominals extend beyond non-canonical argument order, pointing to the additional request to extract information compressed in a dense and more opaque noun phrase to reconstruct the event and integrate it into the unfolding discourse representation (Gaeta, 2002; Ravelli, 1996; Schleppegrell, 2004). The higher probability of partial rescanning with respect to active structures further suggests that readers relied more on rereading when correctly answering questions about nominalized events.

Comprehension accuracy showed a different pattern: passive texts elicited more accurate responses than active texts, whereas nominalized texts did not differ significantly from them. Although passive constructions are known to pose greater comprehension difficulties than actives during language development (Moscati et al., 2023), their advantage here may reflect the specific requirements of the task. Notice that all questions targeted event participants: presenting the patient as the grammatical subject and explicitly introducing the agent through a by-phrase may make the requested referent easier to identify, which was always the focus of the question. The higher accuracy for passives may therefore reflect greater accessibility of the information needed to answer the question.

All in all, our findings highlight the contribution of scanpath analysis: although total dwell time and regressions did not differ significantly across conditions, scanpaths revealed differences in how children moved through and revisited the text. By considering the sequence, location, and duration of fixations, this approach can thus capture aspects of reading behavior that accuracy results and aggregate measures alone cannot describe.

4.2 Processing differences across reader profiles

The second research question addressed whether the processing cost associated with nominalizations varies across developmental

reader groups. As a first step, the background measures provided a clear characterization of the three reader profiles across decoding, receptive vocabulary, grammatical comprehension, and reading comprehension. Children with dyslexia showed marked weaknesses in decoding as well as in reading comprehension, together with lower receptive vocabulary and a tendency toward weaker grammatical comprehension. Bilingual children, who were selected to have no marked decoding impairments, showed preserved word and nonword reading but lower receptive vocabulary and reading comprehension than typically developing monolingual peers. Consistently, the experimental results reveal distinct reading patterns across the three groups in the eye-tracking study too, with typically developing monolingual readers providing a baseline against which the profiles of dyslexic and bilingual readers can be interpreted.

Typically developing monolingual children showed the most efficient and stable processing profile across conditions. The cluster analysis conducted on question answering behavior revealed that minimal rescanning was their most likely reading strategy and was more probable than in the other groups. This suggests that they could draw effectively on information obtained during initial reading to answer comprehension questions, with limited need to return to the text. They also showed lower scanpath dissimilarity than children with dyslexia on both screens, performing similarly to bilingual readers. Nevertheless, it is worth noting that nominalizations triggered higher dissimilarity than both verbal constructions in a similar way across groups, suggesting that they place additional demands on developing readers too, although alongside a generally efficient approach to retrieving the information required by the comprehension question. Overall, these findings indicate that typically developing monolingual readers are sensitive to the processing demands associated with nominalizations but are generally able to manage them efficiently.

Children with dyslexia displayed a clear pattern of increased processing effort across several eye-tracking measures. During initial reading, they exhibited longer total dwell times and higher scanpath dissimilarity than both typically developing groups, together with more regressions than monolingual controls. Longer dwell times and higher dissimilarity relative to monolingual controls were also observed during question-answering. These findings are consistent with the well-documented reading profile associated with dyslexia, characterized by reduced decoding automatization and slower word identification processes (Carioti et al., 2021; Snowling et al., 2020; Vender, 2017). Indeed, even in transparent orthographies such as Italian, dyslexic readers show markedly increased numbers of fixations and longer fixation durations, reflecting a slow but accurate decoding strategy (Hutzler and Wimmer, 2004; De Luca et al., 1999, 2002). Because word identification is slower and more effortful, greater cognitive resources must be allocated to basic decoding processes (Perfetti and Stafura, 2014). Moreover, the weaknesses in lexical access and morphosyntactic competence that characterize dyslexia, as discussed in the Introduction, together with working memory limitations, might have contributed to difficulties in integrating linguistic information during reading to build a coherent representation of the text. The cluster analysis showed that children who answered correctly were more likely to adopt maximal rescanning, with extensive and lengthy re-readings of the text, suggesting difficulty in retrieving information about the event participants.

Importantly, although decoding inefficiency may have contributed to the overall processing cost observed in children with dyslexia, the pattern was not confined to initial reading. The persistence of group differences during question-answering suggests that decoding difficulties may interact with higher-level integration and monitoring processes during reading. As a consequence, fewer resources may remain available for higher-level integration processes, which are required when reconstructing argument structure or in integrating information across sentences. This is consistent with the higher probability of maximal rescanning observed in the dyslexic group relative to monolingual controls, even in correctly answered trials, which suggests increased reliance on rereading when they provided a correct answer. Accuracy in question-answering was also lower than in bilinguals, although no differences with typical monolinguals were observed. Taken together, these findings are consistent with accounts linking difficulties in text comprehension to decoding difficulties, but also of language and processing weaknesses (Araújo et al., 2015; Berninger et al., 2008; Cardinaletti et al., 2022; Georgiou et al., 2023; Melloni and Vender, 2022; Snowling et al., 2020; Vender, 2017).

Importantly, although nominalizations were associated with increased processing demands overall, the analyses did not provide clear evidence of a greater cost in children with dyslexia than in the other groups; instead, their reading behavior was characterized by a generalized increase in processing effort across all conditions. This may indicate broader difficulties in the overall efficiency of the reading system. When decoding processes are slow and resource-demanding, any additional processing demand (such as reconstructing argument structure and thematic relations) may augment the overall cognitive load. These findings are thus consistent with previous research showing that comprehension difficulties in dyslexia arise from the interaction between decoding inefficiency, linguistic knowledge, and limited working memory resources (Georgiou et al., 2022).

A different pattern emerged for ***bilingual readers***, who showed an intermediate profile between monolingual children with dyslexia and those with typical development, particularly in question-answering. During initial reading, no differences in total dwell time or scanpath dissimilarity were detected between bilingual and typically developing monolingual readers. This is consistent with evidence showing relatively small group differences in decoding between monolinguals and bilinguals (Melby-Lervåg and Lervåg, 2014).

Differences emerged more clearly in rereading measures. Compared to monolingual controls, bilingual children produced more regressions when answering comprehension questions, indicating greater reliance on rereading during successful question-answering. The scanpath analysis further clarifies these differences across groups. While dyslexic readers showed more variable and less stable reading strategies, bilingual readers patterned more closely with typically developing monolinguals

in terms of overall scanpath regularity. Bilingual readers' dissimilarity scores did not differ significantly from those of monolingual controls on either screen. However, in correctly answered trials, bilinguals were less likely than monolingual controls to adopt minimal rescanning. Within the bilingual group, minimal and partial rescanning were both more likely than maximal rescanning, with no significant difference between them.

Within the Reading Systems Framework, this pattern may reflect knowledge-based differences in lexical and morphosyntactic representations (Perfetti and Stafura, 2014), rather than process-based inefficiencies. Bilingual children often have less automatized or less richly specified lexical representations in each language (Bialystok et al., 2008, 2010), which may affect word-to-text integration. Critically, empirical evidence indicates that, in bilingual populations, reading comprehension difficulties are primarily driven by weaknesses in language comprehension rather than decoding (Melby-Lervåg and Lervåg, 2014). This distinguishes bilingual readers from dyslexic readers, whose difficulties are strongly linked to decoding inefficiency and reduced processing fluency. This interpretation is consistent with the combination of lower receptive vocabulary and the absence of marked decoding difficulties in our bilingual sample. Nominalizations may place additional costs on processing resources, increasing demands in terms of lexical and morphosyntactic knowledge. Despite differences in reading strategies, comprehension accuracy did not differ significantly from that of typically developing monolinguals, and was higher than that of readers with dyslexia. The combination of preserved accuracy and greater reliance on rereading during question-answering is consistent with a reading profile in which successful comprehension is accompanied by more extensive revisiting of the text. Overall, bilingual readers differ qualitatively from both groups: compared with typically developing monolinguals, they showed comparable early reading measures but greater reliance on rereading during question-answering, despite similar comprehension accuracy. Compared with dyslexic readers, they showed instead less generalized processing difficulty and preserved early processing, although both groups displayed greater reliance on rereading during question-answering than monolingual controls.

4.3 General discussion

The general picture emerging from our results supports the view that nominalizations represent a critical interface between lexical/syntactic processing and discourse integration. As discussed in the Introduction, they compress complex events into nominal units and often obscure the mapping between syntactic position and thematic roles (Gaeta, 2002; Grimshaw, 1990; Schleppegrell, 2004). As a result, readers must reconstruct the underlying event structure in order to interpret the sentence correctly.

From the perspective of the Reading Systems Framework (Perfetti, 2007; Perfetti and Stafura, 2014), nominalizations place simultaneous demands on several components of the reading system. First, readers must access the lexical representation of the nominalized form, which can also require morphological analysis, especially when the form is not fully lexicalized (Plemenitaš, 2015). Second, they must reconstruct the argument structure associated with the underlying verb, identifying the participants involved in the event (Halliday and Martin, 1993). Third, they must integrate this information into the unfolding discourse representation while maintaining previously processed information in working memory. Our aim was to examine how nominalized constructions affect text processing, including the reorganization of the event and its participants within a noun phrase. Accordingly, the observed effects reflect the processing of these constructions as a whole, including the changes they entail in the grammatical organization of the event and its participants. In this spirit, our findings suggest that the success of this process depends on both the efficiency and robustness of the reader's linguistic system. Readers with slower decoding processes (as in dyslexia) may experience generalized processing inefficiencies, while readers with less extensive lexical knowledge in the language of testing may encounter greater difficulty during integration processes. In correctly answered trials, both groups were less likely than monolingual controls to adopt minimal rescanning, indicating that successful question-answering could be accompanied by greater reliance on rereading.

While our study provides new insights into the processing of nominalizations in developing readers, an area on which there is virtually no research (with the exception of Daroczy et al., 2025), several limitations should be acknowledged. First, we focused on a relatively specific age group (fifth- and sixth-grade students). Future research could examine how nominalization processing develops across different stages of reading development. In particular, younger readers may experience greater difficulty in reconstructing argument structure, as found by previous developmental studies (Yatsushiro and Alexiadou, 2020). Second, our study examined a limited set of nominalization types, which was not systematically selected for morphological type or transparency, and the small number of items representing each type prevented us from examining whether converted nominalizations or different types of derived nominalizations are processed differently. As noted in the introduction, nominalizations can vary widely in their morphological transparency and degree of lexicalization. Future research could investigate whether different types of nominalizations impose different processing demands depending on their morphological structure and frequency. Future studies could investigate additional sources of variation in reading patterns, including socioeconomic status, which was not assessed in the present study. They could also examine how individual language experience and similarities between nominalized constructions in children's family languages and Italian influence processing. Given the relatively small number of trials per condition of this work, future studies with larger item sets could provide more precise estimates, particularly for smaller effects and interactions, examining whether the findings observed across the different

measures, including the scanpath analyses, generalize to independent samples. Finally, further work could explore how contextual cues and discourse structure influence the processing of nominalizations. Because nominalized events often rely on contextual information to recover argument structure and temporal relations, providing richer or sparser discourse contexts may accordingly influence processing costs and comprehension outcomes.

5 Conclusions and educational implications

In conclusion, the present study has shown that event nominalizations influence the online dynamics of reading comprehension across different profiles of developing readers. While they do not necessarily impair comprehension accuracy, they increase processing effort and trigger more variable reading strategies during both first reading and question-answering. Moreover, the results reveal distinct processing patterns across reader groups. Children with dyslexia show generalized processing inefficiencies associated with slower decoding and increased rereading, whereas bilingual children display processing patterns that, in our view, reflect differences in linguistic (especially, lexical) representation.

These findings highlight the role of nominalizations as a linguistic structure that places simultaneous demands on multiple components of the reading system. By comparing eye-tracking measures across reader populations, the present study contributes to a deeper understanding of how linguistic complexity interacts with different readers' characteristics in modulating reading comprehension.

Beyond their theoretical significance, our findings have important educational implications. As discussed in the introduction, nominalizations are a hallmark of academic language and are widely used in textbooks and scientific discourse (Biber and Gray, 2016; Hyland and Jiang, 2021). As a result, students must learn to interpret linguistically dense structures in order to access academic content effectively. Our results suggest that nominalizations impose additional processing demands during both initial reading and question-answering even for typically developing readers and may represent a challenge for students with reading vulnerabilities. For dyslexic students, the combination of slow decoding and complex linguistic structures could increase the overall cognitive load during reading; on the other hand, for bilingual students, the linguistic density of nominalizations could place additional demands on lexical knowledge in the language of schooling. Therefore, our findings highlight the potential importance of explicit teaching of nominalizations and, possibly, other complex grammatical constructions often found in scientific and educational texts. In particular, didactic approaches that highlight the relationship between nominalized forms and their underlying verbal structures may help students unpack the semantic content of complex sentences and improve their comprehension of academic texts.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://osf.io/t7ky5/overview.

Ethics statement

The studies involving humans were approved by Comitato di Approvazione della Ricerca sulla Persona (CARP) at the University of Verona (Protocol No. 10.R2/2024). The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation in this study was provided by the participants' legal guardians/next of kin.

Author contributions

MV: Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Writing – original draft, Writing – review & editing. IV: Data curation, Formal analysis, Investigation, Visualization, Writing – original draft, Writing – review & editing. AN: Conceptualization, Data curation, Investigation, Methodology, Writing – original draft, Writing – review & editing. SP: Conceptualization, Methodology, Writing – review & editing. SD: Conceptualization, Methodology, Writing – review & editing. CM: Conceptualization, Investigation, Methodology, Writing – original draft, Writing – review & editing.

Funding

The author(s) declared that financial support was received for this work and/or its publication. This work was funded by the European Union -NextGenerationEU, under the PRIN 2022 PNRR project REAding CompreHension for inclusion (REACH): analyzing and enhancing the comprehension of syntactic and semantic structures in children with linguistic fragilities [project code P2022Y5CWX_002; CUP B53D23030370001].

Acknowledgments

We warmly thank all the children who participated in the study and their families. We are also grateful to the headmasters and teachers of IC12 Golosine in Verona and Fondazione Istituto S. Girolamo Emiliani in Milan, as well as to the director and speech therapists of the UOC Infanzia Adolescenza Famiglia (IAF) and Consultori of ULSS-9 Scaligera, Districts 1–2, Verona. Their collaboration and support were essential to the successful implementation of the research.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that Generative AI was used in the creation of this manuscript. Generative AI was only used to support language editing and stylistic refinement during manuscript preparation. All AI-assisted text was reviewed and verified by the author(s), who take full responsibility for the content of the manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/flang.2026.1912839/full#supplementary-material

References

Aguiar, C., Silva, C. S., Guerra, R., Rodrigues, R. B., Ribeiro, L. A., Pastori, G., et al. (2020). Early interventions tackling inequalities experienced by immigrant, low-income, and Roma children in 8 European countries: a critical overview. *Eur. Early Child Educ. Res J.* 28, 58–76. doi: 10.1080/1350293X.2020.1707363

Araújo, S., Reis, A., Petersson, K. M., and Faísca, L. (2015). Rapid automatized naming and reading performance: a meta-analysis. *J. Educ. Psychol.* 107, 868–883. doi: 10.1037/edu0000006

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

Bello, I. (2016). Cognitive implications of nominalizations in the advancement of scientific discourse. *Int. J. English Stud.* 16, 1–23. doi: 10.6018/ijes/2016/2/262921

Berninger, V. W., Raskind, W., Richards, T., Abbott, R., and Stock, P. (2008). A multidisciplinary approach to understanding developmental dyslexia within working-memory architecture: genotypes, phenotypes, brain, and instruction. *Dev. Neuropsychol.* 33, 707–744. doi: 10.1080/87565640802418662

Bialystok, E., Craik, F. I. M., and Luk, G. (2008). Lexical access in bilinguals: effects of vocabulary size and executive control. *J. Neurolinguistics* 21, 522–538. doi: 10.1016/j.jneuroling.2007.07.001

Bialystok, E., Luk, G., Peets, K. F., and Yang, S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Biling Lang. Cogn.* 13, 525–531. doi: 10.1017/S1366728909990423

Biber, D., and Gray, B. (2016). *Grammatical Complexity in Academic English: Linguistic Change in Writing*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511920776

Biber, D., Johansson, S., Leech, G., Conrad, S., and Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. London: Longman.

Bishop, D. V. M. (2009). "Speech pathology test collection," in *TROG-2: Test for Reception of Grammar—Version 2: Manuale,* eds. S. Suraniti, R. Ferri, and V. Neri, Italian adaptation (Florence: Giunti O.S), 68.

Cardinaletti, A., Piccoli, E., and Volpato, F. (2022). "Dyslexia and syntactic deficits: overview and a case study of language training of relative clauses," in *A Linguistic Approach to the Study of Dyslexia*, eds. G. Cappelli, and S. Noccetti (Bristol: Multilingual Matters), 360. doi: 10.2307/jj.22679767.12

Carioti, D., Masia, M. F., Travellini, S., and Berlingeri, M. (2021). Orthographic depth and developmental dyslexia: a meta-analytic study. *Ann. Dyslexia* 71, 399–438. doi: 10.1007/s11881-021-00226-0

Casalis, S., Colé, P., and Sopo, D. (2004). Morphological awareness in developmental dyslexia. *Ann. Dyslexia* 54, 114–138. doi: 10.1007/s11881-004-0006-z

Cominetti, F. (2023). Nominalization as an enhancer of linguistic implicitness in political discourse. *Lingue e Linguaggi* 56, 69–88. doi: 10.1285/i22390359v56p69

Comrie, B., and Thompson, S. A. (2007). "Lexical nominalization," in *Language Typology and Syntactic Description, 2nd Edn*, eds. T. Shopen (Cambridge: Cambridge University Press), 334–381. doi: 10.1017/CBO9780511618437.006

Cornoldi, C., Colpo, G., and Carretti, B. (2017a). *Prove MT 1-2-3: Kit scuola secondaria di I grado*. Giunti EDU.

Cornoldi, C., Colpo, G., and Carretti, B. (2017b). *Prove MT–Kit scuola: Classi 3-4-5 primaria.* Dalla valutazione degli apprendimenti di lettura e comprensione al potenziamento. Giunti EDU.

Daroczy, G., Artemenko, C., Wolska, M., Meurers, D., and Nuerk, H.-C. (2025). Are text comprehension and calculation processes in word problem solving sequential or interactive? An eye-tracking study in children. *Canadian J. Experim. Psychol.* 79, 206–211. doi: 10.1037/cep0000366

De Cat, C. D., Kašćelan, D., Prévost, P., Serratrice, L., Tuller, L., Unsworth, S., et al. (2023). How to quantify bilingual experience? Findings from a Delphi consensus survey. *Biling Lang. Cogn.* 26, 112–124. doi: 10.1017/S1366728922000359

De Luca, M., Borrelli, M., Judica, A., Spinelli, D., and Zoccolotti, P. (2002). Reading words and pseudowords: an eye movement study of developmental dyslexia. *Brain Lang.* 80, 617–626. doi: 10.1006/brln.2001.2637

De Luca, M., Di Pace, E., Judica, A., Spinelli, D., and Zoccolotti, P. (1999). Eye movement patterns in linguistic and non-linguistic tasks in developmental surface dyslexia. *Neuropsychologia* 37, 1407–1420. doi: 10.1016/S0028-3932(99)00038-X

Downing, A. (2000). "Nominalization and topic management in leads and headlines," in *Discourse and Community: Doing Functional Linguistics,* ed. E. Ventola (Tübingen: Gunter Narr), 355–378.

Dunn, L. M., and Dunn, L. M. (2000). "Encyclopedia of autism spectrum", in *Peabody: Test di Vocabolario Recettivo*, eds. G. Stella, C. Pizzoli, and P. E. Tressoldi, Italian adaptation (Torino: Omega), 1–5.

Elff, M. (2025). *mclogit: Multinomial Logit Models for Categorical Responses and Discrete Choices* (R package version 0.9.15) [Computer software].

Fang, Z., Schleppegrell, M. J., and Cox, B. E. (2006). Understanding the language demands of schooling: nouns in academic registers. *J. Literacy Res.* 38, 247–273. doi: 10.1207/s15548430jlr3803_1

First, M. B. (2013). *DSM-5 Handbook of Differential Diagnosis*. Washington, DC: American Psychiatric Publishing. doi: 10.1176/appi.books.9781585629992

Gaeta, L. (2002). *Quando i Verbi Compaiono Come Nomi: Un Saggio Di Morfologia Naturale*. Milano: FrancoAngeli.

Gamer, M., Lemon, J., Fellows, I., and Singh, P. (2019). *irr: various coefficients of interrater reliability and agreement. R package version 0.84.1*. Available online at: https://CRAN.R-project.org/package=irr

Georgiou, G. K., Martinez, D., Vieira, A. P. A., Antoniuk, A., Romero, S., and Guo, K. (2022). A meta-analytic review of comprehension deficits in students with dyslexia. *Ann. Dyslexia* 72, 204–248. doi: 10.1007/s11881-021-00244-y

Georgiou, G. K., Vieira, A. P. A., Rothou, K. M., Kirby, J. R., Antoniuk, A., Martinez, D., et al. (2023). A meta-analysis of morphological awareness deficits in developmental dyslexia. *Sci. Studies Read.* 27, 253–271. doi: 10.1080/10888438.2022.2155524

Givón, T. (1984). *Syntax: A functional-typological introduction, Vol. 1*. Amsterdam: John Benjamins. doi: 10.1075/z.17

Grabe, W., and Yamashita, J. (2022). *Reading in a Second Language: Moving from Theory to Practice, 2nd Edn*. Cambridge: Cambridge University Press. doi: 10.1017/9781108878944

Grimshaw, J. (1990). *Argument Structure*. Cambridge, MA: MIT Press.

Halliday, M. A. K., and Martin, J. R. (1993). *Writing Science: Literacy and Discursive Power*. London: Routledge.

Hansen, S., Dirksen, R., Küchler, M., Kunz, K., and Neumann, S. (2006). Comprehensible legal texts: Utopia or a question of wording? On processing rephrased German court decisions. *J. Lang. Commun. Business* 19:15. doi: 10.7146/hjlcb.v19i36.25836

Hartig, F. (2026). *DHARMa: Residual Diagnostics for Hierarchical (multi-level/mixed) Regression Models* (R package version 0.5.0) [Computer software]. Available online at: https://CRAN.R-project.org/package=DHARMa

Hutzler, F., and Wimmer, H. (2004). Eye movements of dyslexic children when reading in a regular orthography. *Brain Lang.* 89, 235–242. doi: 10.1016/S0093-934X(03)00401-2

Hyland, K., and Jiang, F. K. (2021). Academic naming: changing patterns of noun use in research writing. *J. Eng. Linguist.* 49, 255–282. doi: 10.1177/00754242211019080

INVALSI (2025). *Rapporto Prove INVALSI, 2025*. Rome: INVALSI.

Iordăchioaia, G., and Melloni, C. (2023). The zero suffix in English and Italian deverbal nouns. *Zeitschrift für Sprachwissenschaft* 42, 109–132. doi: 10.1515/zfs-2022-2014

Koptjevskaja-Tamm, M. (1993). *Nominalizations*. London: Routledge.

Lenth, R. V., and Piaskowski, J. (2025). *emmeans: Estimated Marginal Means, Aka Least-Squares Means* (R package version 2.0.1) [Computer software].

Lyon, G. R., Shaywitz, S. E., and Shaywitz, B. A. (2003). A definition of dyslexia. *Ann. Dyslexia* 53, 1–14. doi: 10.1007/s11881-003-0001-9

Marinelli, C. V., Iaia, M., Zoccolotti, P., Romano, D., Traficante, D., Cassibba, R., et al. (2026). Reading comprehension of children acquiring a transparent language as L2: a study with the simple view of reading model. *Biling Lang. Cogn.* 29:234–248. doi: 10.1017/S1366728925000318

Marinis, T., and Chondrogianni, V. (2010). Production of tense marking in successive bilingual children: when do they converge with their monolingual peers? *Int. J. Speech Lang. Pathol.* 12, 19–28. doi: 10.3109/17549500903434125

Melby-Lervåg, M., and Lervåg, A. (2014). Reading comprehension and its underlying components in second-language learners: a meta-analysis of studies comparing first- and second-language learners. *Psychol. Bull.* 140, 409–433. doi: 10.1037/a0033890

Melloni, C. (2011). *Event and Result Nominals: A Morpho-Semantic Approach*. Bern: Peter Lang. doi: 10.3726/978-3-0351-0335-9

Melloni, C., and Vender, M. (2022). Morphological awareness in developmental dyslexia: playing with nonwords in a morphologically rich language. *PLoS ONE* 17:e0276643. doi: 10.1371/journal.pone.0276643

Mézière, D. C., Yu, L., McArthur, G., Reichle, E. D., and von der Malsburg, T. (2024). Scanpath regularity as an index of reading comprehension. *Sci. Studies Read.* 28, 79–100. doi: 10.1080/10888438.2023.2232063

Moscati, V., Marini, A., and Biondo, N. (2023). What a thousand children tell us about grammatical complexity and working memory: a cross-sectional analysis on the comprehension of clitics and passives in Italian. *Appl. Psycholinguist.* 44, 1161–1184. doi: 10.1017/S0142716423000462

Mueller, B. (2015). *Analysis of nominalization in elementary and middle school science textbooks* [Student capstone thesis, Hamline University]. School of Education and Leadership Student Capstone Theses and Dissertations. Available online at: https://digitalcommons.hamline.edu/hse_all/247

Novita, S., Lockl, K., and Gnambs, T. (2022). Reading comprehension of monolingual and bilingual children in primary school: the role of linguistic abilities and phonological processing skills. *European J. Psychol. Educ.* 37, 993–1013. doi: 10.1007/s10212-021-00587-5

OECD (2023). *PISA 2022 Results (Volume I): The State of Learning and Equity in Education*. Paris: OECD Publishing.

Perfetti, C. (2007). Reading ability: lexical quality to comprehension. *Sci. Studies Read.* 11, 357–383. doi: 10.1080/10888430701530730

Perfetti, C., and Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Sci. Studies Read.* 18, 22–37. doi: 10.1080/10888438.2013.827687

Plemenitaš, K. (2015). "Aspects of nominalization from a cognitive perspective," in *Meaning Making in Text: Multimodal and Multilingual Functional Perspectives*, eds. S. Starc, C. Jones, and A. Maiorani (London: Palgrave Macmillan), 72–91. doi: 10.1057/9781137477309_5

R Core Team (2025). *R: A Language and Environment for Statistical Computing [Computer software]*. Vienna: R Foundation for Statistical Computing.

Ravelli, L. J. (1996). Making language accessible: successful text writing for museum visitors. *Linguist. Educ.* 8, 367–387. doi: 10.1016/S0898-5898(96)90017-0

Sartori, G., Job, R., and Tressoldi, P. E. (2007). *DDE-2: Batteria Per La Valutazione Della Dislessia E Della Disortografia Evolutiva-2*. Florence: Giunti O.S.

Schleppegrell, M. J. (2004). *The Language of Schooling: A Functional Linguistics Perspective*. Mahwah, NJ: Routledge. doi: 10.4324/9781410610317

Scrucca, L., Fraley, C., Murphy, T. B., and Raftery, A. E. (2023). *Model-Based Clustering, Classification, and Density Estimation using Mclust in R*. Boca Raton, FL: Chapman and Hall/CRC. doi: 10.1201/9781003277965

Snowling, M. J., Hulme, C., and Nation, K. (2020). Defining and understanding dyslexia: past, present and future. *Oxf. Rev. Educ.* 46, 501–513. doi: 10.1080/03054985.2020.1765756

Stepanov, A., Andreetta, S., and Pavlič, M. (2026). Heritage language in a border-shift context: differential input effects on sentence comprehension. *Biling Lang. Cogn.* 25, 1–13. doi: 10.1017/S1366728925100904

Venables, W. N., and Ripley, B. D. (2002). *Modern Applied Statistics with S, 4th Edn*. New York, NY: Springer. doi: 10.1007/978-0-387-21706-2

Vender, M. (2017). *Disentangling Dyslexia: Phonological and Processing Impairment in Developmental Dyslexia*. Bern: Peter Lang. doi: 10.3726/b11503

Vender, M., Delfitto, D., and Melloni, C. (2018). Clitic production in bilingual children: when exposure matters. *Languages* 3:22. doi: 10.3390/languages3030022

von der Malsburg, T. (2018). *scanpath: Tools for Analyzing Spatio-Temporal Patterns in Eye Movements* (R package version 1.06).

Wolfer, S. (2016). "The impact of nominalisations on the reading process: a case-study using the Freiburg Legalese Reading Corpus," in *Eyetracking and applied linguistics*, eds. S. Hansen-Schirra, and S. Grucza (Berlin: Language Science Press), 163–186.

Yang, L., Li, C., Li, X., Zhai, M., An, Q., Zhang, Y., et al. (2022). Prevalence of developmental dyslexia in primary school children: a systematic review and meta-analysis. *Brain Sci.* 12:240. doi: 10.3390/brainsci12020240

Yatsushiro, K., and Alexiadou, A. (2020). "The acquisition of argument-roles in nominalizations," in *BUCLD 44: Proceedings of the 44th annual Boston University conference on language development*, eds. M. M. Brown, and A. Kohut (Somerville, MA: Cascadilla Press), 749–758.

Ziegler, J. C., and Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: a psycholinguistic grain size theory. *Psychol. Bull.* 131, 3–29. doi: 10.1037/0033-2909.131.1.3