UNIVERSITÀ DEGLI STUDI DI VERONA

DEPARTMENT OF FOREIGN LANGUAGES AND LITERATURES
GRADUATE SCHOOL OF HUMANITIES
DOCTORAL PROGRAM IN LINGUISTICS

In agreement with the
Free University of Bozen–Bolzano

XXXVIII CYCLE / 2022

Prediction in Spoken Language Comprehension across Linguistic Levels: Evidence from L2 Speakers and Adults with Dyslexia

S.S.D. L-LIN/01

Coordinator: Prof. Stefan RABANUS (University of Verona)
Signature ____________

Tutor: Prof. Lorenzo SPREAFICO (Free University of Bozen–Bolzano)
Signature ____________

Tutor: Prof.ssa Maria VENDER (University of Verona)
Signature ____________

Doctoral Student: Natalya SHIROKORAD
Signature ____________

Abstract

Language comprehension is not only incremental but predictive: over the past two decades, extensive evidence has shown that comprehenders do not passively wait for linguistic input but anticipate words and structures before they are encountered, thereby facilitating efficient comprehension. While it is no longer surprising that prediction exists, important questions remain about how it operates across different linguistic levels and population groups. Research on second-language (L2) speakers has produced conflicting findings, with some studies demonstrating predictive use of semantic, morphosyntactic, or phonological cues, while others report reduced or even absent anticipatory processing. Even less is known about prediction in developmental dyslexia – very few studies to date have examined whether individuals with dyslexia can anticipate upcoming linguistic input, leaving this population strikingly under-represented in the literature. Yet understanding prediction in dyslexia is crucial, as the condition is characterised by persistent difficulties in phonological and morphological processing, slower lexical retrieval, and differences in verbal working memory. These factors may plausibly affect anticipatory processing not only in reading but also in spoken language comprehension.

This dissertation addresses these gaps by investigating predictive mechanisms at the semantic, morphosyntactic, and phonological levels, comparing neurotypical L1 adults, i.e., adults without self-reported reading or language impairments, with L2 speakers and adults with dyslexia. Two visual-world eye-tracking studies were conducted. Experiment 1 (administered in Italian) comprised two parts: one testing anticipation of a highly expectable target (contextual prediction) and its phonological form, while the other examined the predictive use of grammatical gender cues on the determiner. It was administered to native speakers of Tyrolean who acquired Italian after the age of 3, alongside neurotypical L1 Italian adults and adults with dyslexia. Experiment 2 (administered in English) was a close replication of Ito and Husband (2017), focusing on contextual, semantic, and phonological prediction, and extended their study by including second-language speakers and adults with dyslexia. In contrast to Experiment 1, it also included a semantic competitor condition to test whether semantically related alternatives of highly predictable targets receive anticipatory activation.

Results revealed a consistent ability across all groups to anticipate a single highly predictable target word. Beyond this, the strength and timing of prediction varied across linguistic levels and participant groups. In Experiment 1, L2 participants showed somewhat shallower target prediction with similar onset timing to L1 controls, while individuals with dyslexia displayed broadly preserved prediction. Although none of the groups displayed anticipatory looks driven by grammatical gender cues, neurotypical L1 and L2 readers showed a gender facilitation effect after word onset, which was not observed in the dyslexia group. Importantly, there was no evidence for phonological pre-activation in any group. In Experiment 2, all groups anticipated the highly predictable target, but the semantic competitor condition revealed selective spreading activation: only competitors that shared semantic features with both the target and other words in the preceding context attracted anticipatory looks, whereas competitors related only to the target did not. Adults with dyslexia showed significantly delayed initiation of competitor looks, while L2 speakers exhibited only a trend towards slower semantic pre-activation. Again, no phonological competitor effects emerged in predictable sentences.

These findings indicate that spreading activation from highly predictable targets to semantically related but contextually implausible words is limited and constrained by contextual relevance. The parallel absence of both

semantic and phonological competitor effects in highly predictable contexts suggests that strong contextual support inhibits cohort competitors rather than reflecting a general absence of phonological form pre-activation. This pattern points to an efficient cognitive architecture that appears to strike a balance between the benefits of anticipation and the costs of activating contextually irrelevant alternatives.

Contents

List of Tables

List of Figures

General Overview

This section provides a concise roadmap to the dissertation, outlining its theoretical motivation, research aims, and organizational structure. Understanding how humans comprehend language is one of the central questions in psycholinguistics. A key component of this process is prediction, which is broadly defined as the pre-activation of upcoming linguistic information, i.e., words, forms, or structures, before it is encountered (Altmann & Mirković, 2009; Huettig, 2015; Kuperberg & Jaeger, 2016; Pickering & Gambi, 2018). Prediction is therefore understood as an anticipatory mechanism that draws on information available in the unfolding sentence to facilitate real-time comprehension. Over the past couple of decades, research has largely moved beyond asking *whether* prediction exists to exploring *how* it operates across linguistic levels (e.g., semantic, morphosyntactic, phonological) and *who* uses it most efficiently.

There are several critical questions in the prediction literature that remain unresolved. First, evidence regarding phonological form prediction is mixed: some studies have reported anticipatory activation of phonological form of the target before its acoustic onset (DeLong et al., 2005; Ito et al., 2018, 2020; Li & Qu, 2023; Li et al., 2022), whereas others found no reliable pre-activation even in highly constraining contexts (Ito & Husband, 2017; Ito & Sakai, 2021; Nieuwland et al., 2018). Second, evidence for semantic competitor pre-activation is limited, with only a small set of studies addressing whether listeners pre-activate concepts related to, but distinct from, the most predictable word in context (Dijkgraaf et al., 2019; Ito & Husband, 2017; Li & Qu, 2023; Li et al., 2022). Third, research on prediction in developmental dyslexia is strikingly limited. Existing work suggests delayed or less efficient anticipatory processing in moderately constraining contexts, where upcoming words are plausible but not highly expected, and reduced use of grammatical cues (Engelhardt et al., 2021; Huettig & Brouwer, 2015), while phonological prediction in dyslexia has not yet been systematically examined. This population is particularly informative because dyslexia involves persistent difficulties in phonological awareness, slower lexical and verbal processing (Georgiou et al., 2018; Hayiou-Thomas et al., 2018; Lyon et al., 2003; Snowling et al., 2020), all of which may weaken predictive processing. Finally, prediction in second language (L2) speakers often depends on proficiency, exposure, and cue type (Kaan et al., 2016; Pickering & Gambi, 2018; Tagliani et al., 2025), making L2 comprehension informative for examining how linguistic experience shapes anticipatory mechanisms. Moreover, evidence suggests that L2 speakers may weight linguistic cues differently than L1 speakers due to crosslinguistic influence and differences in cue reliability across their languages (Kaan, 2014; Schlenter, 2023), leading to qualitatively different prediction patterns. For instance, predictive cues that are absent or differently encoded in the L1, i.e., grammatical gender, case marking, or morphophonological agreement, may be less readily exploited during L2 processing (Dussias et al., 2013; Hopp, 2013; Hopp & Lemmerth, 2018; Ito et al., 2024; Lew-Williams & Fernald, 2010; Mitsugi & Macwhinney, 2016). Similarly, reduced reading experience in individuals with dyslexia has been associated with delays in anticipatory processing, though the mechanisms underlying these delays remain debated (Engelhardt et al., 2021; Huettig & Brouwer, 2015). Taken together, adults with dyslexia and L2 speakers were selected for this study because each group highlights different limitations on predictive processing. While dyslexia provides a test case for examining whether phonological and reading-related difficulties constrain prediction, L2 speakers enable investigation of how reduced exposure, variable proficiency, and cross-linguistic influence modulate anticipatory processing. By studying both populations, this dissertation addresses whether prediction is a universal, automatic mechanism or a resource-sensitive process that depends on individual cognitive profiles and language experience.

In this dissertation, I aim to contribute to the current research by first providing new insights into how prediction mechanisms function in adults with developmental dyslexia and in L2 speakers during spoken language comprehension; second, I address current debates about the linguistic levels at which prediction occurs by examining semantic, morphosyntactic, and phonological prediction, and by testing whether these levels are modulated by L2 status or by reading-related difficulties (with a focus on developmental dyslexia). Before presenting the empirical work, **Chapter 1** provides the theoretical background. It reviews models of spoken word recognition and sentence processing, outlines current approaches to predictive processing, and summarises the linguistic characteristics of the populations studied. The empirical work consists of two visual-world eye-tracking studies. Both experiments used the visual-world paradigm, in which participants listened to sentences while viewing arrays of pictures, and their eye movements were tracked to assess whether they anticipated upcoming linguistic input in real time.

Experiment 1 (Italian), presented across **Chapters 2** and **3**, was carried out at the Free University of Bozen–Bolzano, Italy, with L1 Italian typical readers, L1 Italian readers with dyslexia, and L2 Italian speakers with Tyrolean as their L1[1]. The experiment consisted of two parts, administered to the same participants within the same session:

- *Experiment 1a*[2] (Chapter 2) tested contextual prediction (anticipation of a specific highly predictable lexical item based on sentence-level constraints) and phonological-form prediction (pre-activation of the sound form of contextually predicted word before its acoustic onset).
- *Experiment 1b* (Chapter 3) examined morphosyntactic prediction,

[1]The participants in this study are also bilinguals in the sense that they regularly use both languages in everyday life. However, I use the term "L2 speakers" throughout to emphasize the acquisition-based distinction (Italian learned after L1 Tyrolean, typically in formal educational settings) rather than simultaneous bilingual acquisition.

[2]Because Experiment 1 has two parts with the same participants run in a single session, I refer to them as *1a* and *1b* rather than as separate experiments.

specifically whether anticipatory eye movements are guided by the grammatical gender information encoded in Italian determiners. The chapter also considers the potential influence of cross-linguistic gender congruency between Italian and Tyrolean in L2 speakers.

Experiment 2 (English) presented in **Chapter** 4, was conducted at the University of East Anglia, UK, and closely replicated and extended Ito and Husband (2017). Three participant groups were tested: L1 English typical readers, L1 English readers with dyslexia, and L2 English speakers with Romance L1s (Italian, Spanish, French, Portuguese, or Romanian). Like Experiment 1, it examined contextual prediction, phonological-form prediction, and additionally included semantic-competitor pre-activation (pre-activation of items semantically related to the highly predictable target). English does not have a system of nominal gender agreement on determiners; therefore, morphosyntactic prediction via grammatical gender was not examined in this experiment.

Although the experiments differed in their implementation, both studies were designed to investigate the same questions: whether listeners anticipate upcoming linguistic input, and whether this ability varies across populations and linguistic levels. Finally, **Chapter** 5 offers a general discussion integrating findings across experiments, theoretical implications, and outlines the study's limitations along with directions for future research. The **Appendices** provide experimental materials, statistical outputs, and correlation analyses.

Chapter 1

Spoken Language Comprehension and Predictive Processing: From Recognition to Anticipation

1.1 Theoretical Foundations of Spoken Word Recognition and Sentence Processing

Understanding spoken language is an extraordinary cognitive feat. From a stream of rapidly unfolding sounds, proficient listeners are able to identify words, map them onto stored knowledge, and build a coherent interpretation of sentences and discourse. This ability is far from trivial: unlike written language, which clearly separates words with spaces, the speech signal is continuous, highly variable across speakers and contexts, and lacks clear boundaries between words (see overviews in Dahan and Magnuson 2006; Huettig 2015). Yet human listeners achieve recognition with remarkable speed and accuracy.

Over the past several decades, researchers across linguistics, psychology, neuroscience, and computer science have been trying to unravel the mechanisms behind this phenomenal capability of humans by developing a range of theoretical models to explain how this is possible. These approaches differ in the weight they assign to bottom-up processes, which start with the acoustic signal and build upward toward higher-level representations[1], and top-down processes, in which prior knowledge and context shape perception from the start. Bottom-up approaches emphasize the role of the input itself, while top-down approaches stress the influence of expectations, lexical knowledge, and contextual cues. Most contemporary models adopt an interactive perspective, assuming bidirectional information flow between perceptual and lexical levels, whereby information can move both bottom-up (from lower to higher processing levels) and top-down (from higher to lower levels)(see Jackendoff, 2007; Magnuson et al., 2018; Mcclelland et al., 2006; McClelland & Elman, 1986).

The following sections trace the evolution of these ideas. I begin with traditional, largely bottom-up accounts of speech perception and spoken word recognition, then turn to models emphasizing interactivity between multiple sources of information. Finally, I review models of sentence processing, where similar debates about modular versus constraint-based architectures occur, shedding light on our understanding of how words are combined into meaningful structures.

1.1.1 Spoken Word Recognition: From Bottom-Up to Interactive Models

To comprehend speech, listeners must retrieve words from their mental lexicon, the store of long-term knowledge about word forms and meanings, and map the acoustic signal onto discrete linguistic units that can be

[1]In cognitive and psycholinguistic models, representations refer to mental encodings of linguistic information at different levels of abstraction. They can be conceived as stored or activated patterns corresponding to phonological forms, lexical entries, semantic features, or conceptual knowledge, organized within networks in which activation can spread between related nodes (Collins & Loftus, 1975; Dell, 1986; Levelt, 1989). "Lower-level" representations correspond to more perceptual and form-based information such as phonetic or phonological detail, while "higher-level" representations correspond to more abstract structures such as syntactic, semantic, or discourse-level information that integrate meaning and context.

identified as words. However, comprehension does not end there: listeners must also access the meanings of these words and integrate them into higher-level syntactic and semantic representations. Spoken word recognition therefore provides a bridge between low-level perception and higher-order cognitive operations of retrieval, parsing, and interpretation (Dahan & Magnuson, 2006, p. 249).

Early approaches to language processing were often conceived as serial stage models, in which comprehension and production were divided into distinct steps occurring one at a time (Fromkin, 1971). Applied to speech perception, this meant that recognition was assumed to proceed in a strictly bottom-up fashion: processing began with the raw sound signal, which was incrementally analyzed into phonetic segments (such as [p] or [a]), then assembled into larger phonological forms, and only then recognized as words with meaning. Each level of processing was thought to be completed before the next began, with little or no influence from higher-level knowledge (Carroll, 2008; D. Pisoni & Levi, 2005; D. B. Pisoni & Levi, 2012).

However, bottom-up accounts that view speech as a linear string face several challenges. First, co-articulation, or context-conditioned variability, means that the pronunciation of a sound depends heavily on its phonetic environment: the /t/ in *ten* sounds different from the /t/ in *better* or *intact*. Second, the problem of speech segmentation arises because the acoustic signal is continuous, with few reliable breaks between words. To a listener unfamiliar with a language, the input can sound like an undifferentiated stream. Moreover, restricting models to low-level perceptual decoding leaves unexplained how listeners access lexical representations, select among competitors, and integrate recognized words into syntactic and semantic structures to construct coherent meaning. For example, many spoken words are phonemically embedded within others (for example, *part*, *art*, *men*, and *ment* within *apartment*), making it impossible to determine word boundaries based solely on a discrete sequence of phonemes. Successful recognition therefore requires sensitivity to fine-grained subphonemic cues, such as subtle durational and coarticulatory differences, that signal whether a syllable belongs to an independent word or forms part of a longer one (Dahan & Magnuson, 2006). These characteristics make it difficult to imagine how speech could be segmented and decoded purely from the acoustic input alone, without the combined influence of subphonemic detail, contextual information, and higher-level linguistic constraints.

In response, researchers began to emphasize the role of top-down and interactive processes. These are mechanisms by which higher-level information, such as context, word frequency, or semantic plausibility, guides recognition from the outset. Rather than treating perception and comprehension as isolated stages, interactive accounts propose continuous bidirectional information flow between acoustic, lexical, and structural levels of representation (Dahan & Magnuson, 2006). For example, listeners may identify a word's meaning before all of its sounds have been processed, using contextual cues to fill in the gaps (Balota, 1990; D. Pisoni & Levi, 2005). In this sense, top-down mechanisms complement bottom-up analysis: the acoustic signal provides accuracy, whereas contextual and linguistic knowledge ensure efficiency and flexibility in dealing with variability and ambiguity. Thus, modern theories generally assume that word recognition involves some balance of both.

In what follows, I distinguish between two broad classes of models that capture this historical and conceptual shift. The Cohort Model (Marslen-Wilson & Welsh, 1978) and the Neighborhood Activation Model (NAM) (Luce & Pisoni, 1998) represent primarily bottom-up accounts in which lexical candidates are activated and compete as the speech signal unfolds, though competition itself introduces dynamic selection processes beyond simple feed-forward decoding. Interactive and connectionist models, in contrast, extend these principles by allowing bidirectional information flow and parallel activation across multiple representational levels (Magnuson et al., 2018; Mcclelland et al., 2006; McClelland & Elman, 1986). This division highlights the field's evolution from relatively autonomous bottom-up processing with lexical competition toward fully interactive frameworks of spoken word recognition.

Cohort and Neighborhood Models. In the late 20th century, Marslen-Wilson and Welsh (1978) proposed one of the first influential alternatives to purely serial, feed-forward accounts - **the Cohort Model**. While still largely bottom-up in architecture, it proposed that word recognition unfolds incrementally as the speech signal arrives, with activation and competition among lexical candidates driving comprehension beyond simple phoneme identification. The model assumes that when a listener hears the beginning of a word, it activates a "cohort" of lexical candidates that share the same onset. As the acoustic signal unfolds, candidates that no longer match are gradually eliminated until only one remains, at which point recognition occurs. A central concept is the uniqueness point - the moment when the input provides enough information to identify a word uniquely in the lexicon. For example, the word *trespass* can only be uniquely identified at the /p/ sound, where it diverges from competitors such as *tress* or *trestle*; in context, however, recognition may occur earlier if these competitors are syntactically or semantically excluded (see Marslen-Wilson, 1987). Evidence for the Cohort Model came from gating tasks (Grosjean, 1980), where listeners are presented with increasingly larger audio fragments of a word and asked to guess the word after each fragment. Results showed that recognition often occurred before the entire word was heard, consistent with the idea of incremental narrowing. Subsequent work built on the Cohort Model to examine whether contextual information could constrain lexical activation from the earliest moments of processing. Some studies, such as Grosjean (1980), found evidence for contextual preselection, suggesting that semantic or discourse cues can restrict the initial cohort (i.e., the set of lexical candidates that share the same onset and are temporarily activated when a word begins) even before full bottom-up information becomes available. In contrast, Tyler (1984) argued that the cohort is generated purely bottom-up and that context exerts its influence only later, by suppressing candidates that are inconsistent with the developing interpretation. These findings opened an important line of research on the timing of contextual effects, foreshadowing later debates about interactivity in spoken-word recognition.

A related but distinct account is **the Neighborhood Activation Model (NAM)** (Luce & Pisoni, 1998). Whereas the Cohort Model focuses on words that overlap at onset, NAM emphasizes that recognition involves competition among all words that are phonologically similar to the target, regardless of position. It formalizes lexical competition statistically, proposing that recognition speed and accuracy depend on both the frequency of the target word and the density of its phonological neighborhood. Words with many high-frequency neighbors are harder to recognize than words with sparse neighborhoods because competition among candidates is stronger.

Both the Cohort Model and NAM played a crucial role in shifting attention away from strictly serial, bottom-up accounts. Although both remain primarily bottom-up in architecture, they introduced key dynamic principles – incremental activation, competition, and graded lexical access – that challenged the view of recognition as a simple left-to-right decoding process. The two models differ in their assumptions: the Cohort Model describes recognition as the progressive elimination of onset-based candidates, while NAM proposes continuous competition among all phonologically similar neighbors. Despite these differences, both highlight the dynamic nature of word recognition, paving the way for frameworks that would explicitly model bidirectional information flow between levels of representation (Mcclelland et al., 2006; McQueen et al., 2003).

Interactive and Connectionist Models. Building on the insights of the Cohort and Neighborhood Activation Models, later approaches emphasized that multiple levels of representation can be activated in parallel and influence each other dynamically. These are often referred to as interactive or connectionist models.

The most influential of these is the **TRACE model** (McClelland & Elman, 1986), which implemented the interactive activation framework computationally to see how well it accounts for psychological data on human speech perception. In TRACE, different layers of processing units represent acoustic–phonetic features, phonemes, and words. There is constant interaction between the levels of representation. Activation spreads upward from features to phonemes and words, while feedback allows higher levels to influence lower levels, inhibiting competing word units until only one word is left active, which is the word recognized by the system. Empirical support for these dynamics came from the visual world eye-tracking study by Allopenna et al. (1998). Participants heard instructions such as "Pick up the beaker" while viewing a display with four objects: the target (beaker), a cohort competitor (beetle), a rhyme competitor (speaker), and an unrelated distractor (carriage). Eye movements showed that listeners briefly looked not only at the cohort competitor but also at the rhyme competitor, even though it overlapped with the target only at the end. This was the first demonstration of rhyme activation during spoken word recognition, indicating that lexical candidates do not need to share onsets to become active. Such findings are well explained by TRACE, in which activation can spread throughout the network and allow multiple candidates to compete dynamically as the signal unfolds.

Other models further developed the idea of continuous, overlapping activation. In **continuous cascaded models**, language processing happens in a cascade - information flows forward to the next processing stage before the earlier stage has fully completed, allowing partial activation to influence higher-level representations, e.g., from the prelexical level of word-form representations to the word meanings (McQueen et al., 2003). This view contrasts with strictly serial accounts, in which each stage must finish before the next begins. Cascade models provide a more flexible account of incremental recognition, where listeners can begin integrating information about meaning before all of a word's sounds have been processed (McQueen et al., 2003). Interestingly, the cascade principle also applies to picture recognition, where the flow of information is reversed: visual features activate semantic representations, which in turn can influence phonological codes (Huettig & McQueen, 2007).

A broader theoretical perspective on interactivity within language processing is offered by the **Parallel Architecture (PA)** proposed by

Jackendoff (2007). While the models discussed above focus primarily on the mechanisms of spoken word recognition, PA provides a more general framework for understanding how linguistic information is represented and integrated across multiple levels of analysis. It assumes that phonology, syntax, and semantics constitute independent but interconnected systems, each governed by its own principles and linked through interface rules that allow information to flow bidirectionally between levels. This organization contrasts with traditional generative models, which treated syntax as the single generative core from which sound and meaning were derived. Figure 1.1 illustrates the Parallel Architecture framework, in which the three generative components, phonology, syntax, and semantics, are connected by interface mappings that support the dynamic exchange of information across levels during language processing.

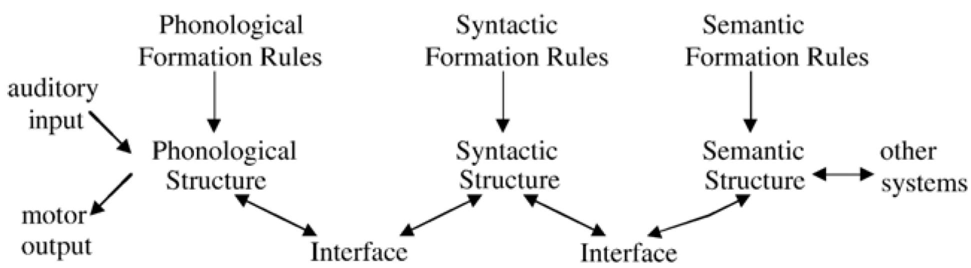


Figure 1.1: Schematic illustration of the Parallel Architecture framework. *Note.* From Jackendoff, R. (2007). A Parallel Architecture perspective on language processing. *Brain Research, 1146*, 2–22. © 2007 Elsevier B.V. Reproduced with permission from Elsevier.

Together, these frameworks capture the mechanisms of word recognition and the interplay among representational levels. A separate line of research has focused on how listeners build syntactic and semantic structures at the sentence level during comprehension.

1.1.2 Models of Sentence Processing

Beyond word recognition, comprehension also requires listeners to construct and revise sentence structures in real time. Models of sentence processing explain how people resolve structural ambiguity, integrate multiple sources of information, and arrive at coherent interpretations as speech unfolds. This is central to understanding the balance between serial and parallel processing and the role of context in shaping comprehension.

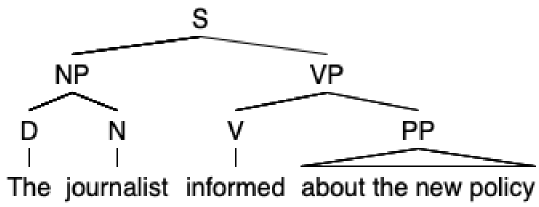
To parse a sentence means to determine its syntactic structure by breaking it down into constituents, such as nouns, verbs, and modifiers, and establishing how they are related to each other. This process of how we comprehend language seems to be effortless, but some sentence structures are harder to process than others. Because language is inherently ambiguous, comprehenders do not always arrive at the intended interpretation immediately. The cost of this ambiguity is that the parser can make mistakes, sometimes committing to an analysis that later proves incorrect. A well-known example of such temporary misanalysis is the garden-path effect. Consider the following sentence fragments:

(1a) The journalist informed about the new policy...

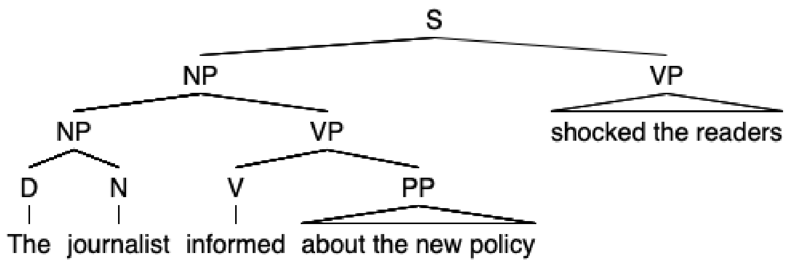
(1b) ...shocked the readers.

(1c) The journalist who was informed about the new policy shocked the readers.

In (1a), the first part of the sentence is structurally ambiguous: the verb *informed* can initially be misinterpreted as the main verb of the sentence. Processing difficulty arises when the continuation (1b) is encountered, forcing a reanalysis in which *informed about the new policy* is understood as part of a reduced relative clause. By contrast, in (1c) the relative pronoun *who* removes the ambiguity and no reanalysis is required. This processing difficulty is due to the less frequent syntactic interpretation (reduced relative clause vs. main clause). Figures 1.2a and 1.2b illustrate the two competing syntactic structures. Figure 1.2a shows the initial, garden-path analysis, where *informed* is parsed as the main verb. Figure 1.2b shows the reanalyzed structure, in which *shocked the readers* is correctly identified as the main verb phrase, and *informed about the new policy* is parsed as part of a relative clause. Early theories proposed that such garden-path effects arise



(a) Initial parse (garden-path analysis).



(b) Revised parse (reanalysis with relative clause).

Figure 1.2: Alternative syntactic structures for the sentence "The journalist who was informed about the new policy shocked the readers." (a) The initial parse misinterprets *informed* as the main verb. (b) The revised parse identifies *shocked the readers* as the main verb phrase, with *informed about the new policy* forming a relative clause.

because the sentence processor is modular and serial. According to the classic Garden Path theory (Frazier, 1979), the parser initially constructs a single syntactic structure using minimal syntactic information, guided by principles such as Minimal Attachment and Late Closure, which favor the simplest and most local syntactic structure possible. In this two-stage model, syntactic processing occurs first and autonomously, and semantic or contextual information can influence comprehension only later, during reanalysis. Thus, according to this approach, reduced relative clauses, such as "The journalist informed about the new policy..." are initially misinterpreted because the parser favors the simplest structural attachment before considering plausibility or context.

Constraint-based models (MacDonald et al., 1994; Tanenhaus & Trueswell, 1995; Trueswell et al., 1994), in contrast to serial syntax-first models, assume that multiple syntactic interpretations are activated in parallel, and that their probabilities are continuously updated by weighted constraints. These constraints derive from lexical frequency, thematic plausibility, prosodic and discourse factors, and even non-linguistic information such as visual context (Pickering & Gompel, 2006). In this view, sentence processing is not modular and syntax-first, but rather interactive and probabilistic from the outset. Parsing decisions reflect graded competition among alternatives, and comprehension emerges as the most highly weighted analysis wins.

Already in the late 1990s and early 2000s, a series of visual-world eye-tracking studies[2] demonstrated that syntactic and semantic interpretation are continuously influenced by the surrounding visual or situational context (Spivey, 2002; Tanenhaus & Trueswell, 1995). A classic demonstration of this principle is the eye-tracking study by Tanenhaus and Trueswell (1995). In this paradigm, participants hear spoken sentences while viewing a visual scene, and their eye movements to objects are tracked. Because eye movements are time-locked to the unfolding speech, anticipatory looks provide a sensitive index of prediction before the relevant word is heard. In their study, participants heard temporarily ambiguous instructions such as "Put the apple on the towel in the box" while viewing a display containing objects arranged in different ways. The prepositional phrase *on the towel* could be interpreted as either a current location of the apple (NP-attachment) or a destination (VP-attachment). In one version of the display, when only one apple was present (Figure 1.3, left), listeners often misinterpreted on the towel as a destination, consistent with a garden-path analysis. But when two apples were present (one on a towel, one on a napkin, see Figure 1.3, right), eye-movement patterns showed that listeners immediately used the visual context to interpret on the towel as a modifier, correctly identifying which apple was meant. These findings demonstrate that comprehenders integrate linguistic input with visual

[2] The visual world paradigm (VWP) is a method in which participants' eye movements are tracked as they listen to spoken sentences while viewing a visual display. Because eye movements are time-locked to the unfolding speech signal, they provide a sensitive, real-time measure of linguistic processing and anticipatory prediction (Cooper, 1974)

information in real-time to guide syntactic decisions, challenging the idea of an encapsulated syntactic module.

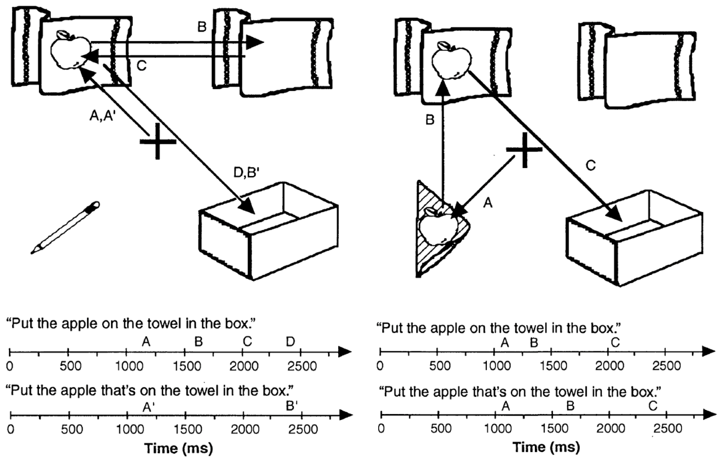


Figure 1.3: Experimental displays and eye-movement patterns. *Note.* From Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*(5217), 1632–1634. © 1995 American Association for the Advancement of Science. Reproduced with permission from AAAS.

Subsequent work extended this principle to even richer situational constraints. In two experiments, Chambers et al. (2004) tracked listeners' eye movements while they followed instructions with temporary syntactic ambiguities, such as "Pour the egg in the bowl over the flour." The affordances of candidate referent objects were manipulated: sometimes both eggs in the display were liquid and could be poured, while other times only one egg was in liquid form (see Figure 1.4). The results showed that syntactic decisions depended on whether multiple action-compatible referents were available. When only one egg could be poured, listeners were "garden-pathed" into misinterpreting *in the bowl* as a goal phrase. But when two pourable eggs were present, this misinterpretation disappeared, as listeners instead treated *in the bowl* as a modifier to identify the intended egg. According to these findings parsing decisions are guided not only by linguistic information but also by situation-specific affordances relevant to achieving the described action.

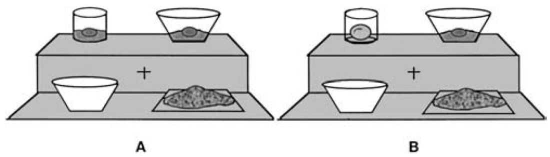


Figure 1.4: Experimental displays illustrating the manipulation of action affordances in syntactic ambiguity resolution. *Note.* From Chambers, C. G., Tanenhaus, M. K., & Magnuson, J. S. (2004). Actions and affordances in syntactic ambiguity resolution. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(3), 687–696. © 2004 American Psychological Association. Reproduced with permission.

Overall, these visual-world studies provided compelling real-time evidence against strongly modular accounts of sentence processing. Their findings illustrate that comprehension involves the continuous integration of multiple constraints, including lexical, semantic, visual, and pragmatic, from the earliest moments of processing, indicating that language comprehension cannot be explained by an autonomous, syntax-first mechanism but instead involves dynamic interaction among linguistic and non-linguistic representations.

While constraint-based accounts emphasize the integration of multiple probabilistic cues, other perspectives argue that comprehenders do not always construct fully detailed syntactic representations and sometimes rely on less detailed analyses. According to the **"good enough"** approach, parsing may rely on surface-level or heuristic interpretations that suffice for communication, even if they do not fully resolve structural ambiguities (Ferreira et al., 2002). This approach explains why listeners sometimes accept semantically plausible but structurally incorrect interpretations (e.g., "The dog was bitten by the man"), suggesting that comprehension balances efficiency with accuracy.

As shown above, constraint-based models of sentence processing have been highly influential in demonstrating that linguistic interpretation is incremental and immediately sensitive to visual context. These accounts establish that visual information can act as a powerful constraint on interpretation from the earliest moments of processing. At the same time, much of this work has focused on object-based reference resolution and spatial relations, leaving more open questions about how information about depicted events and actions is selected and integrated as the utterance unfolds in time.

A broader theoretical perspective that builds on these insights is offered by **the Coordinated Interplay Account (CIA)** proposed by Knoeferle and Crocker (2006, 2007). While constraint-based models emphasize the weighting of multiple probabilistic cues when alternative interpretations compete, the CIA specifies a mechanism by which linguistic input dynamically coordinates visual attention, allowing information from the scene to constrain interpretation precisely when it becomes linguistically relevant. Moreover, visual context is not assumed to be uniformly available at all times; instead, the unfolding utterance (particularly verbs) licenses which aspects of the scene become informative for interpretation. Conceptually, the CIA assumes that different aspects of comprehension operate in parallel and overlap in time, rather than unfolding in a strict step-by-step sequence. These include (i) sentence interpretation, which builds meaning and generates expectations from linguistic input, (ii) utterance-mediated allocation of visual attention to relevant scene elements, and (iii) scene integration, which reconciles linguistic interpretations with visual information. The account assumes bidirectional information flow: linguistic input guides visual attention (as when hearing "the apple" directs gaze toward depicted apples), while visual context can inform and revise linguistic interpretation (as when depicted events help resolve syntactic ambiguities before linguistic cues fully disambiguate).

To further incorporate variability across comprehenders and contexts, Münster and Knoeferle (2018) extended the CIA into the **social Coordinated Interplay Account (sCIA)**. The sCIA captures both biological characteristics (e.g., age, working memory) and experiential characteristics (e.g., literacy, language proficiency, cultural knowledge), as well as socially interpreted contextual cues such as speaker voice, facial expressions, or gaze. These factors are assumed to modulate the strength and timing with which linguistic and visual information interact during comprehension. Thus, the sCIA provides a unifying framework for interpreting existing findings. For example, evidence that emotional state modulates the strength of expectancy-based processing (Van Berkum et al., 2013) can be understood as reflecting differences in how comprehender characteristics shape the coordination between linguistic input and contextual information. Similarly, socially driven expectation violations, such as increased processing difficulty when "Every evening I drink some wine" is spoken in a child's voice (Van Berkum et al., 2008), illustrate how socially interpreted cues can rapidly interact with linguistic expectations. By integrating comprehender characteristics and social context into a real-time processing framework, the sCIA provides theoretical grounding for examining how individual differences in reading ability, cognitive resources, and language experience may modulate spoken language processing.

The theoretical landscape surveyed in this section reveals a broad convergence toward interactive, constraint-based accounts of spoken language comprehension. Across both word recognition and sentence processing, the evidence favors frameworks in which multiple sources of information, such as acoustic, lexical, syntactic, semantic, and contextual, are integrated in parallel and in real time, rather than being processed serially or in encapsulated stages. Several theoretical frameworks are of particular relevance to the present work. At the word recognition level, the TRACE model and the Parallel Architecture provide the representational foundations for understanding how activation spreads across phonological, lexical, and semantic levels, and how this spreading activation can give rise to anticipatory effects. At the sentence level, constraint-based models and, more specifically, the Coordinated Interplay Account and its social extension (sCIA) establish that comprehension is immediately and continuously shaped by visual context, individual characteristics, and social information, providing the theoretical grounding for examining how populations with different linguistic profiles engage in predictive processing. Taken together, these frameworks motivate a view of comprehension as a fundamentally predictive process: one in which prior knowledge and context pre-activate upcoming linguistic representations across multiple levels. The following section turns directly to this question, examining what prediction is, why and how it occurs.

1.2 Prediction in Language Processing: General Perspectives

In recent decades, prediction has become a central concept not only in psycholinguistics but across the cognitive sciences more broadly. The idea that the human brain functions as a "predictive machine" has been applied to perception, action, memory, and decision-making (Bar, 2007; Clark, 2013; Friston, 2010). Within this view, the mind continuously generates hypotheses about future states of the environment and updates these hypotheses based on sensory input. Language processing provides a particularly rich framework for these ideas because comprehension unfolds incrementally over time, offering many opportunities to observe how prior knowledge shapes expectations about upcoming material. Contextual constraint is typically indexed via cloze probability (CP) - the proportion of participants in a norming study who produce a given continuation to a sentence (Taylor, 1953). Higher cloze is associated with processing facilitation on expected words, reflected in faster behavioral responses (e.g., shorter reading or listening times) and in reduced neural responses to semantic processing difficulty (DeLong et al., 2005; Kutas & Federmeier, 2011; Kutas & Hillyard, 1984; Rayner & Well, 1996; Staub et al., 2015; Wlotko & Federmeier, 2012). Thus, most of the empirical literature on prediction is based on this operational link between contextual expectancy and facilitation.

1.2.1 What is Prediction?

Building on the incremental and interactive view of language processing outlined above, prediction is typically operationalized as the pre-activation of linguistic representations, ranging from general semantic features to syntactic structures and fine-grained phonological forms, prior to the arrival of corresponding bottom-up input (Kuperberg & Jaeger, 2016; Pickering & Gambi, 2018; Van Petten & Luka, 2012).

A commonly drawn contrast is with priming. In its classic formulation, priming refers to facilitation arising from prior exposure, often via automatic spreading activation from a presented item to related items (e.g., the word *salt* activating *pepper*) (Neely & Kahan, 2001) and is typically described as fast and automatic. On this view, prediction is a context-driven mechanism that generates expectancies at the level of discourse message, whereas priming is a stimulus-triggered associative phenomenon (Otten & Van Berkum, 2008). Pickering and Gambi (2018), for instance, explicitly treat classic spreading-activation priming as a form of "prediction-by-association" (e.g., *king* pre-activating *queen/thing*) while distinguishing it from more structured routes that can yield more specific and longer-lived pre-activation ("prediction-by-production"), further discussed in section 1.2.3.

However, as Huettig (2025) argues, this dichotomy is often overstated. He notes that researchers sometimes appeal to a heuristic contrast that priming is short-lived "implicit memory" and prediction is longer-lasting, deliberate expectancy, even though the empirical record contains many counterexamples, including long-lasting priming effects. These "fussy" distinctions make it difficult to determine whether specific facilitatory effects should be attributed to one or the other. Huettig (2025) therefore recommends an inclusive definition that treats prediction as "the conscious or subconscious use of information from previous experiences for the conscious or subconscious processing of information about future states of the body and environment" (p. 16). This inclusive stance does not deny that some facilitation occurs at the level of integration (i.e., only when the target arrives is incoming material combined with context for easier comprehension). Rather, it highlights that, in many cases, priming and contextual prediction are hard to disentangle and may even reflect different stages or routes of a single process. Pickering and Gambi (2018) also do not exclude that spreading activation ("prediction-by-association") may constitute the initial semantic phase of "prediction-by-production", consistent with the idea that production and comprehension share a semantic network.

In the present work, I follow Huettig (2025) and adopt an inclusive definition of prediction: any prospective use of prior experience to process future input. This definition encompasses both automatic, short-lived effects and deliberate, context-driven expectations, unified by their forward-looking pre-activation of upcoming linguistic representations. Empirically, facilitation can occur both before and after the relevant input. However, only the former – effects that happen prior to or at the onset of the expected material, which can be detected through anticipatory eye movements, early neural responses, or other time-sensitive measures – provides strong evidence of prediction (Pickering & Gambi, 2018). From this perspective, prediction in language processing is best seen as a graded phenomenon ranging from automatic associative activation to more deliberate, context-based expectations. Subsequent sections will explore why prediction occurs and how predictions are generated.

1.2.2 Why Do We Predict?

A long-standing question is why comprehenders engage in prediction at all. Early accounts suggested that generating predictions might be risky because any erroneous guesses would incur recovery costs, known as "prediction error". According to this view, prediction could only be beneficial in contexts where one continuation was nearly certain; otherwise, disconfirmed predictions would impose a higher cognitive burden than simple integration of the bottom-up input as it arrived (Forster, 1981).

More recent work, however, presents two complementary perspectives, typically examined by assessing whether less predictable continuations impose additional processing costs. One emphasizes *facilitation*: predictions enhance efficiency and speed by preparing the system for likely continuations, such that even when expectations are not met, processing proceeds without slowdown (e.g., Brothers et al., 2023; Frisson et al., 2017; Luke & Christianson, 2016). The other emphasizes *learning and adaptation*: when disconfirmed expectations elicit measurable costs, these prediction errors are thought to provide valuable feedback that helps update the system's internal models of language (e.g., Bannon et al., 2025).

Evidence for facilitation without additional costs comes from both corpus and experimental studies. Luke and Christianson (2016) linked cloze probabilities of words in a large corpus of naturally occurring texts to eye-movement patterns during reading. They found that only a small proportion of content words were highly predictable (around 5% of content words), but when they were, such words were read faster and received fewer fixations. Crucially, when readers encountered a less-expected but still plausible word, no slowdown was observed, suggesting that prediction enhances efficiency without necessarily imposing costs when expectations are not met.

Frisson et al. (2017) extended this line of work with a controlled eye-tracking experiment using carefully constructed sentences. They compared highly constraining contexts that strongly supported one expected completion but also allowed a plausible alternative, such as "For Halloween, Liz dressed up as an ugly old witch/ghost and then went to the party," where *witch* is highly predictable and *ghost* is probable but less expected. The same critical words were also embedded in weakly constraining (neutral) contexts (e.g., "For her arts class, the little girl drew a witch/ghost and then showed it to everyone"). The results of their first experiment demonstrated that unpredictable but plausible words (e.g., *ghost*) were read just as quickly in constraining as in neutral contexts, showing no evidence of a prediction-error cost. Readers, therefore, did not commit to a single lexical candidate but rather appeared to maintain graded activation of multiple plausible continuations, consistent with interactive, probabilistic models of comprehension. In a second experiment, the authors added a semantic-relatedness manipulation, introducing a third, unrelated alternative (e.g., *nurse* in "For Halloween, Liz dressed up as an ugly old witch/ghost/nurse and was very excited"). Here, semantically related but low-cloze words (like *ghost*) were processed more easily in constraining than in neutral contexts, eliciting fewer regressions and shorter total reading times, whereas no such facilitation occurred for unrelated words (such as *nurse*). This facilitation could reflect several possibilities: (a) pre-activation of semantically related words through spreading activation from the predicted target, (b) independent contextual activation of a set of words sharing relevant semantic features, or (c) easier post-lexical integration of semantically related words into the discourse context, especially since the effect emerged only in relatively late eye-movement measures. Thus, while the first experiment provides clear evidence for graded pre-activation without processing cost, the nature of the related-word facilitation remains unclear. Supporting this conclusion, Brothers et al. (2023) used event-related potentials (ERPs) to test whether pre-activating multiple candidates creates interference or facilitation. Their results showed that semantically related alternatives enhanced rather than inhibited each other's activation, further confirming that prediction can involve the parallel activation of multiple candidates. In this way, prediction appears to implement the same principles of graded competition and weighting described in broader models of sentence and word processing in the previous Section, 1.1.

Other work, however, highlights the costs of failed predictions. Bannon et al. (2025) contrasted unexpected but semantically related continuations with anomalous ones. In highly constraining contexts, participants sometimes encountered (a) a semantically related but unexpected word (e.g., *skull* in "The frontal lobe is an important part of the skull," where *brain* was expected), or (b) an anomalous continuation that violated the context entirely (e.g., *brain* in "In the story, Snow White ate a poisoned brain," instead of the expected *apple*). In low-constraining contexts, multiple continuations were plausible and no single prediction was dominant. Results showed that semantically related but low-probability words (e.g., *skull*) were processed more slowly than words in neutral contexts, and anomalous continuations imposed even greater costs (e.g., *brain*). Thus, prediction errors might not always be harmless: even plausible but unexpected continuations can slow down processing. The authors interpret these costs not as wasted effort but as signals that support learning by enabling the language learning system to adjust expectations and adapt and reduce future prediction errors.

Although much of the empirical evidence reviewed above comes from reading, the distinction between facilitation and learning through prediction error should not be restricted to that modality. In spoken language comprehension, there is evidence that prediction errors can serve as learning signals. For example, developmental work by Reuter et al. (2019) demonstrated that children's erroneous predictions during spoken word learning facilitated the acquisition of novel labels. Similarly, Bovolenta and Marsden (2024) showed that in adult L2 learners, encountering verbs in unexpected syntactic contexts during auditory artificial language learning promoted stronger adaptation and generalization of novel structures, consistent with error-based learning accounts.

In later chapters, I will return to spoken-language paradigms such as the visual world and ERP to show how these mechanisms operate in real time. For now, these studies highlight that both perspectives - prediction as facilitation and prediction as learning - are compatible, and indeed complementary. Prediction increases efficiency when input matches expectations, but when it does not, errors themselves provide useful information for adaptation. Moreover, both perspectives challenge earlier models of prediction that viewed it as an all-or-nothing phenomenon, where erroneous predictions were seen as a burden.

1.2.3 How Are Predictions Generated?

The automaticity of prediction in language processing remains a subject of ongoing research. Evidence suggests that prediction involves a combination of both automatic and resource-dependent processes (Ito & Pickering, 2021). Several theories have been proposed to explain how predictions are generated (Huettig, 2015; Pickering & Gambi, 2018; Pickering & Garrod, 2013). Broadly, two mechanisms have been highlighted: prediction-by-production and prediction-by-association. These accounts differ not only in the processes they propose, but also in the extent to which prediction is assumed to be automatic.

According to the *prediction-by-production* account, comprehenders covertly engage their production system to simulate likely upcoming utterances and subsequently compare the outputs of these predictions with the ones of the utterance (Chang et al., 2006; Dell & Chang, 2014; Pickering & Gambi, 2018; Pickering & Garrod, 2013). Supporting this view, neurocognitive evidence shows extensive overlap between production and comprehension neural circuits (Giglio et al., 2022; Scott & Johnsrude, 2003; Walenski et al., 2019). Moreover, Schomers et al. (2015) demonstrated that stimulation of the articulatory motor cortex can causally influence spoken word comprehension, while Scott et al. (2009) and Watkins et al. (2003) found that listening to speech, but not nonspeech, activates tongue and lip muscles. Although neurocognitive evidence demonstrates a substantial overlap between production and comprehension systems, this overlap alone does not prove that prediction depends on production. It suggests an interweaving of perception and production mechanisms, yet experimental evidence is required to show that predictive processing specifically draws on production resources.

Direct support for this view comes from an EEG study by Martin et al. (2018), who went one step further by showing that such motor involvement in comprehension can, at least partly, be explained by its major role in prediction. The authors asked whether blocking access to the production system would diminish prediction during reading. To test this, Spanish speakers were divided into three groups and asked to read highly constraining sentences while performing secondary tasks. In the syllable-production group, participants repeatedly uttered a syllable to prevent covert production; in the tongue-tapping group, participants tapped their tongue without speaking; and in the listening group, participants heard recordings of their own voice producing a syllable. Prediction was measured by assessing the magnitude of the N400 effect[3] elicited by expected versus unexpected gender-marked articles. Only the articulatory-suppression condition reduced the N400 expectation effect, whereas tongue-tapping and listening did not. This pattern indicates that when access to subvocal production is prevented, anticipatory activation of upcoming words is weakened, suggesting that production mechanisms are functionally involved in prediction. However, this effect may instead reflect interference with lexical access processes that are fundamental to comprehension and therefore for predictive language processing, rather than from production mechanisms playing a direct causal role in prediction. At the same time, evidence from early language development indicates that production abilities themselves can enhance prediction. Mani and Huettig (2012) found that two-year-olds with larger productive vocabularies showed stronger anticipatory eye movements during sentence comprehension. Importantly, no correlation between predictive skills and children's comprehension skills was observed. This suggests that access to the production system facilitates predictive processing from an early age.

In addition to the more resource-intensive message-level mechanisms that rely on richly contextual representations integrating syntactic, semantic, and pragmatic information, comprehenders can also anticipate language input through an associative route. This *prediction-by-association* account emphasizes spreading activation between related representations - whether semantic/associative (meaning-based) or phonological (sound-based), as discussed by Pickering and Gambi (2018). In this process, predictions do not rely on message-level discourse context but arise from immediate information stored in short-term memory. This lower-level information triggers activation of related representations, making upcoming words easier to process. Essentially, hearing or thinking about a word can prime the system for related continuations (Forster, 1981). For example, hearing *cat*

[3] The N400 is a negative-going event-related potential (ERP) component peaking around 400 ms after word onset, sensitive to the ease of semantic processing and contextual integration. More negative amplitudes reflect greater difficulty in integrating a word into the context (Kutas & Hillyard, 1980).

activates its representation, which then spreads to related concepts such as *dog* (semantic) or *mat* (phonological).

A visual-world study by Kukona et al. (2011) illustrates how such associative mechanisms can generate predictions beyond strict contextual constraints. Participants heard sentences such as "Toby arrests the crook" while viewing displays containing semantically related objects. Predictive looks occurred not only to the target (*crook*) but also to a semantically related competitor (*policeman*), even though the agent role was already filled and the competitor was contextually inappropriate. This pattern suggests that spreading activation automatically pre-activates multiple related candidates in parallel, regardless of their fit with the discourse.

While the interactive and constraint-based models described in Section 1.1.2 focus on the processes by which information from multiple sources is integrated during comprehension, the Parallel Architecture (PA) originally proposed by Jackendoff (2007) (see Section 1.1.1) and later extended to prediction by Huettig et al. (2022) provides a more general representational framework for explaining how such interactivity is possible. In contrast to models like TRACE or constraint-based approaches that simulate online processing, PA specifies how linguistic knowledge itself is represented in the mind. It assumes that phonological, syntactic, and semantic representations are generated independently but are linked through bidirectional interface mappings, allowing information to flow across levels. This architecture thus offers a principled way to conceptualize how activation of one level (e.g., a semantic concept) can pre-activate corresponding syntactic or phonological structures, giving rise to predictive effects as a natural consequence of representational interconnectivity. A clear visual illustration of this principle is offered by Huettig et al. (2022), who demonstrate how activation within the PA framework can spread both vertically across interface links (e.g., from the sound of a word to its meaning) and horizontally within a representational domain (e.g., through semantic or associative relations). As shown in Figure 1.5, hearing the word *chair* activates its phonological (/tʃɛr/) and semantic representations, which in turn spread activation to a conceptually related item such as sit and its corresponding phonological form. This example captures how predictive activation arises naturally within the representational architecture posited by PA, where associative connections can automatically pre-activate multiple candidates in parallel.

Unlike prediction-by-production, this route does not involve covertly simulating the speaker's utterance. Instead, it operates largely automatically, driven by associations between related representations stored in memory. Prediction-by-association is often considered a non-optional component of comprehension, and predictions are not always constrained by the context. For example, as previously mentioned, Kukona et al. (2011) found anticipatory looks to semantically related competitors even when the sentence context already made such continuations implausible. Such findings highlight the automatic and sometimes non-selective nature of associative prediction.

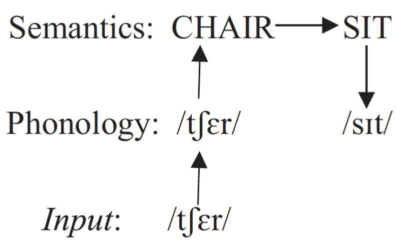


Figure 1.5: Spreading activation in the Parallel Architecture framework. Vertical arrows indicate activation spreading across interface mappings (between phonology and semantics), while horizontal arrows indicate activation within a representational domain (e.g., between semantically related concepts). *Note.* Reproduced from Huettig, F., Audring, J., & Jackendoff, R. (2022). A parallel architecture perspective on pre-activation and prediction in language processing. *Cognition, 222,* Article 105012, under the terms of the Creative Commons CC-BY license.

These two perspectives are often contrasted, but they need not be mutually exclusive: which route dominates likely depends on contextual support and available processing resources. For instance, Boudewyn et al. (2015) showed that listeners flexibly adjust their expectations when encountering unexpected input, which reflects the dynamic interplay between associative activation and higher-level contextual simulation. At the same time, the coexistence of automatic associative activation and more resource-dependent, context-driven expectations highlights an open challenge for current theories, which have yet to specify precisely whether and how these mechanisms interact during comprehension.

1.3 Prediction Across Levels of Linguistic Representation

1.3.1 Contextual and Semantic Prediction

Research on prediction at the semantic level can be divided into two related but distinct questions: (i) whether comprehenders pre-activate highly predictable target words in context (contextual/conceptual prediction), and (ii) whether this activation extends to semantic competitors of those targets. For example, in a sentence such as "On her birthday, Emma blew out the…", listeners may strongly anticipate the word *candles*, demonstrating contextual prediction. The question is whether semantically related concepts, such as *matches*, *lamp*, or *flame* are also pre-activated, reflecting broader semantic spreading independent of contextual constraints. While the former is widely attested and largely uncontroversial, the latter has received comparatively less attention and remains an open question. In this subsection, I first review classic demonstrations of contextual prediction, before turning to the smaller body of work investigating semantic competitor activation.

Classic demonstrations of contextual prediction. A classic demonstration comes from Altmann and Kamide (1999), who were among the earliest to show that verbs with strong selectional restrictions can trigger anticipatory eye movements using a visual world paradigm. For instance, in a scene containing a boy, a toy car, and a cake, hearing "The boy will eat..." led listeners to fixate the cake (the only edible object) well before it was named (see Figure 1.6 for a picture display example). Later studies confirmed and extended this finding, showing anticipatory looks to objects licensed by highly predictable context (e.g., Arai & Keller, 2013; Borovsky et al., 2012; Ito et al., 2018; Kamide et al., 2003).

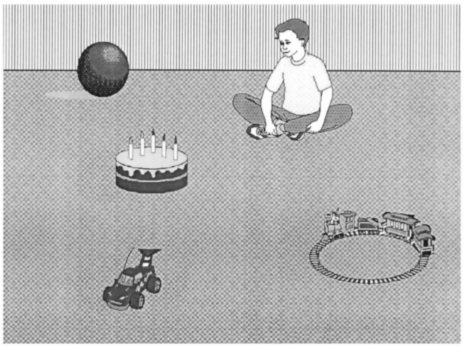


Figure 1.6: Example of a visual scene. The image shows various objects, including a cake, a toy car, a ball, and a toy train. *Note.* From Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73*(3), 247–264. © 1999 Elsevier. Reproduced with permission.

Converging evidence comes from ERP and reading studies. In ERP research, high-cloze targets elicit reduced N400 amplitudes relative to low-cloze or anomalous words, reflecting facilitated access for predicted items (Federmeier, 2007; Kutas & Federmeier, 2011; Van Berkum et al., 2005; Wicha et al., 2004). In reading, predictable words are fixated less often and for shorter durations than unpredictable ones, supporting the view that prediction facilitates lexical access (Rayner, 1998; Staub & Clifton, 2006).

Taken together, these findings leave little doubt that comprehenders routinely anticipate high-cloze target words in strongly constraining contexts. The question is whether this predictive activation also spreads to semantically related alternatives.

Semantic competitor prediction. Whereas contextual prediction of highly expected target words is well established, evidence for the anticipatory activation of their semantic competitors is more limited. Only a handful of studies have directly examined whether comprehenders pre-activate concepts that are semantically related to the most predictable word in context. Importantly, these studies suggest that prediction does extend beyond single target items to include semantically related concepts.

Classic ERP evidence comes from Federmeier and Kutas (1999), who showed that during reading in highly constraining sentence contexts, unexpected words from the same semantic category as the expected completion (e.g., *pines* when *palms* was predicted) elicited reduced N400 amplitudes compared to unrelated unexpected words (e.g., *tulips*). According to this finding, semantic category features of predictable words can be pre-activated before input onset, facilitating the processing of semantically related but unpredicted items. One concern, however, is whether such effects truly reflect anticipatory activation of competitors or instead arise from lexical priming by the individual context words. Addressing this issue, Ito et al. (2016) replicated the N400 reduction for semantically related competitors (e.g., *page* when *book* was predicted), but showed that the effect occurred only in high-cloze sentences. Because the reduction depended on cloze probability rather than plausibility, it is best explained as pre-activation of semantic features of the expected target word, rather than integration ease or lexical priming.

Evidence from spoken language comprehension has been mostly obtained using the visual world paradigm. In a landmark study, Ito and Husband (2017) tracked participants' eye movements as they listened to sentences such as "That dog looks so happy, wagging its tail...". In each trial, participants viewed the visual display 3000 ms prior to sentence onset, allowing sufficient time to encode the scene. Critically, well before the target word *tail* was heard, listeners showed anticipatory looks not only to the target picture (*tail*), but also to a semantically related competitor (*paw*). Time-course analyses revealed that fixations to target objects diverged from unrelated objects as early as ∼4500 ms before target onset, while semantic competitors showed increased fixation proportions from ∼2000 ms before target onset in predictable sentences. These effects were observed across varied semantic relationships, including taxonomic categories (e.g., *nose–ear*), thematic associations (e.g., *tea–lemon*), and part–whole relations (e.g., *car–tire*). In contrast, in unpredictable contexts, these effects emerged only after target onset (∼500 ms post-onset for semantic competitors), indicating that they reflect anticipatory processing rather than bottom-up integration. Extending this work, studies in Mandarin Chinese by Li et al. (2022) and Li and Qu (2023) replicated anticipatory looks to semantic competitors alongside highly predictable targets (these studies are discussed in more detail later in the section 1.3.3 in relation to phonological prediction).

Together, these findings suggest that while semantic competitor prediction has been investigated far less extensively than target prediction, the available evidence consistently indicates activation of semantically related alternatives. However, it remains unclear whether this activation reflects spreading from an anticipated target word, or rather, semantic priming elicited by the context itself, or some combination of both.

1.3.2 Prediction of Morphosyntactic and Syntactic Features

Prediction in language comprehension is not limited to semantic content. A large body of research has examined whether comprehenders also anticipate grammatical features of upcoming words, such as gender, number, or syntactic structure. These studies are of particular interest because they investigate prediction at a morphosyntactic level, often using pre-nominal elements (articles, adjectives) as diagnostic cues. Evidence comes from both event-related potential and eye-tracking paradigms.

ERP evidence from gender-marked cues. Early evidence comes from Spanish. In a seminal study, Wicha et al. (2004) showed that readers generated expectations about the grammatical gender of highly predictable nouns. Participants read high-cloze sentences in which the expected noun was preceded by either a gender-consistent or a gender-inconsistent article. Unexpected articles that mismatched the gender of the most predictable completion (e.g., in Spanish, the masculine determiner *un* instead of the feminine *una* before the feminine noun *corona*) elicited a late positivity around 500–700 ms[4]. This effect on the article itself provides strong evidence

[4]Wicha et al. (2004) interpreted this late positivity as reflecting a violation of gender-based expectations. Late positivities in similar time windows have been associated

that comprehenders pre-activate grammatical gender information of expected words.

Similar findings were reported in spoken comprehension in Dutch by Otten et al. (2007), who showed that, in predictive contexts, gender-mismatching adjectives evoked an early right-frontal negativity (300–600 ms). The authors interpreted this neural response as reflecting the rapid detection of a mismatch between the predicted and the grammatical gender of the adjective, rather than a canonical N400 effect. This effect was absent in control conditions where the context was not predictive, indicating that the gender-based anticipation was driven by discourse-level expectations rather than simple lexical priming. In a follow-up study, Otten and Van Berkum (2009) replicated these findings while examining the role of individual differences in working-memory capacity (WMC). Unlike their earlier experiment, which tested spoken sentences with gender-marked adjectives (Otten et al., 2007), this study used written Dutch stories in which gender-marked determiners preceded the predictable noun. Participants were classified as high- or low-WMC readers based on a standardized Reading Span Test (Daneman & Carpenter, 1983). Both groups read predictive and non-predictive stories while ERPs were recorded at the determiner position. As in their earlier work, unexpected determiners elicited an early right-frontal negativity (300–600 ms) in predictive contexts, indicating that both high- and low-WMC readers anticipated the grammatical gender of the upcoming noun. However, only low-WMC readers showed an additional late negativity (900–1,500 ms) to prediction-inconsistent determiners, suggesting that while WMC does not determine whether predictions are generated, it does modulate how readers process information that disconfirms those predictions. Subsequent ERP studies further confirmed these anticipatory effects. For example, Foucart et al. (2014) reported anticipatory gender effects on prenominal articles in Spanish, and Ito et al. (2020) found that Italian articles elicited robust negativities (250–800 ms) when their gender mismatched the predicted noun, suggesting rapid grammatical feature pre-activation. It is

with the P600 component, which is typically linked to syntactic reanalysis, reprocessing, or detection of agreement violations (Hagoort, 2003; Osterhout & Holcomb, 1992).

worth noting, however, that not all studies have replicated these effects. A large-sample replication by Kochari and Flecken (2019), using nearly identical Dutch materials and procedures to those of Otten and Van Berkum (2009), failed to find significant ERP differences between gender-matching and gender-mismatching articles, although similar patterns were observed, with the same general tendency for unexpected articles to elicit more negative-going responses and comparable scalp distributions. The authors attributed this to methodological factors and to the relatively weak and variable nature of gender marking in Dutch, which may make it a less reliable cue for lexical prediction. Dutch gender marking provides only a weak and sometimes ambiguous cue: the common-gender article *de* is also used for all plural nouns, and the neuter article *het* for all diminutives, regardless of grammatical gender. As a result, encountering an unexpected article (e.g., *de* instead of *het*) does not necessarily force revision of the predicted noun's meaning but may simply lead comprehenders to reinterpret its number or form (e.g., expecting a plural or diminutive variant). According to the authors, this limited distinctiveness of Dutch gender marking likely makes it a less reliable cue for lexical prediction.

Eye-tracking and spoken language evidence. While ERP studies focus on mismatch effects, eye-tracking in the visual world paradigm can directly test whether listeners use grammatical cues to anticipate upcoming referents. Huettig and Brouwer (2015) showed that Dutch adults exploited gender-marked determiners (*de/het*) to anticipate nouns: upon hearing the article, participants shifted gaze toward the gender-matching object in the display before the noun was spoken. Similarly, Bosch and Foppolo (2022) found that Italian-speaking children anticipated upcoming nouns based on article gender. Unlike ERP paradigms, which infer pre-activation of predictable targets indirectly from the processing cost of gender mismatches, eye-movement studies demonstrate more directly that listeners immediately exploit gender-marked cues to guide their gaze toward the correct referent in real time, given the constraints of the visual scene.

Beyond gender: syntactic structure. Prediction also extends beyond morphosyntactic agreement features to more abstract properties of syntactic structure and argument roles. Using the visual world paradigm, Chambers et al. (2004) showed that object affordances constrain the interpretation of prepositional phrases in spoken instructions, guiding anticipatory eye movements (Figure 1.4). Arai and Keller (2013) demonstrated that listeners anticipate argument structure based on verb subcategorization. In their experiments, participants heard sentences containing either transitive or intransitive verbs while viewing a scene with several potential referents. Eye movements revealed that as soon as participants heard the verb, they shifted attention toward an appropriate postverbal argument before it was mentioned. For example, upon hearing "The nun punished...", participants looked at the potential object (*the artist*) prior to its acoustic onset, indicating that they predicted a postverbal direct object. In contrast, when the verb was intransitive, such as "The nun disagreed. . . ," listeners did not show anticipatory looks to any object, since such verbs do not take a direct object. Anticipatory eye movements only appeared once additional material, such as a prepositional phrase (". . . with the artist"), made a new argument structure possible.

This pattern suggests that listeners use verb-specific subcategorization information to generate predictions about upcoming syntactic structure. Such findings receive support from *surprisal theory* (Hale, 2001; R. Levy, 2008), a computational framework proposing that processing difficulty at each word is proportional to its surprisal — the negative log-probability of that word given its preceding context. Under this account, the comprehension system continuously tracks the probability of different possible interpretations of the sentence as it unfolds word by word. When a word is unexpected, more mental work is required to update those expectations — and that is what drives processing difficulty. Importantly, surprisal captures predictability arising from any source (semantic, lexical, or syntactic) since any structural or lexical pattern that affects how probable a word is will automatically influence how easy it is to process (Hale, 2001;

R. Levy, 2008). This suggests that the anticipatory behavior observed in eye-tracking studies reflects, at least in part, the comprehension system's sensitivity to the probabilistic structure of the language itself. Taken together, findings from ERP, eye-tracking, and computational modeling demonstrate that prediction in comprehension extends beyond semantic content to morphosyntactic and structural levels.

1.3.3 Phonological Prediction

While semantic and syntactic prediction has been highly attested, evidence for the prediction of phonological information remains inconclusive, though a growing body of work has examined it using both ERP and visual-world eye-tracking paradigms. By phonological prediction, I refer to the pre-activation of sound-based representations of highly predictable targets.

ERP evidence. One of the first pieces of evidence for phonological prediction comes from DeLong et al. (2005), who tested the prediction of the word-initial phoneme of a highly expected word, making reference to the English article rule that *a* precedes consonant-initial words and *an* precedes vowel-initial words. Participants read highly constraining sentences (e.g., "The day was breezy so the boy went outside to fly ...") where the most likely continuation was *a kite* (cloze $\approx$ 89 %), but other plausible continuations such as *an airplane* were also possible, though less expected. DeLong and colleagues observed that N400 amplitudes on both the expected noun (*kite*) and its preceding article (*a*) correlated with sentence cloze probabilities, with effects observed in the canonical N400 time window (approximately 200–500 ms post word onset) – the more unexpected the article was in the context, the more negative ERP amplitude was observed. The graded N400 effect on the article was interpreted as evidence that comprehenders not only anticipated the upcoming noun conceptually but also pre-activated its phonological form, specifically whether it would begin with a vowel or a consonant.

However, subsequent replications have yielded mixed results. Using a similar ERP paradigm to DeLong et al. (2005), Martin et al. (2013) compared native English speakers and late Spanish–English learners and found article-related N400 differences only in L1 speakers, suggesting that L2 readers did not anticipate the phonological form of the highly predictable word to the same extent. In contrast, a large-scale multi-lab replication across nine laboratories by Nieuwland et al. (2018) directly reproduced DeLong et al.'s (2005) materials and procedures with a larger sample of native English speakers (N = 334). While they successfully replicated the graded N400 effect on the target noun, they failed to reproduce the article-level effect, concluding that phonological pre-activation, if present at all, is considerably weaker and less consistent than initially reported. They attributed this discrepancy to several factors, including differences in statistical power, analysis pipelines, and the small size of the original effect. More recently, Ito et al. (2020) used Italian indefinite articles (*un/una/uno*), which are marked both for initial sound and grammatical gender of the following noun. They reported distinct ERP negativities for phonological and gender mismatches, with gender mismatch effects emerging earlier (250–800 ms) than phonological mismatch effects (450–800 ms), suggesting that phonological prediction is possible but less strong than morphosyntactic prediction.

Visual-world eye-tracking evidence. Another line of research has examined phonological prediction in spoken comprehension using the visual world paradigm. In these studies, participants listen to predictive sentences while viewing displays that include the target object, phonological competitors, and unrelated distractors. Anticipatory fixations to phonological competitors, if present, are taken as evidence that comprehenders pre-activate the phonological form of an expected word.

Ito et al. (2018) conducted a comprehensive study that tested both L1 English speakers and Japanese learners of English (L2), and included semantic as well as phonological competitors. In the phonological condition, participants heard constraining sentences such as "The tourists expected rain when the sun went behind the ...", in a scene containing a target (*cloud*), a phonological competitor (*clown*), and unrelated distractors (see Figure 1.7). The visual display was presented 1000 ms prior to the onset of the critical word. Native participants showed strong anticipatory looks to the target as expected. Importantly, they also showed a short-lived increase in fixations to phonological competitors between 500 and 350 ms before target onset, with fixation proportions to phonological competitors briefly exceeding those to unrelated distractors within this window. This effect was considerably smaller in magnitude than the anticipatory looks to the target and did not persist beyond this narrow time window, suggesting that phonological form was pre-activated, but only transiently (see Figure 1.8, where the blue line represents the proportion of looks to the phonological competitor). This temporal profile indicates that phonological prediction is relatively weak and short-lived: listeners briefly activate the sound form of the expected word (*cloud*), as evidenced by brief fixations on phonological competitors (*clown*), but this activation quickly decays. I will return to their findings regarding L2 participants in Section 1.4.1.

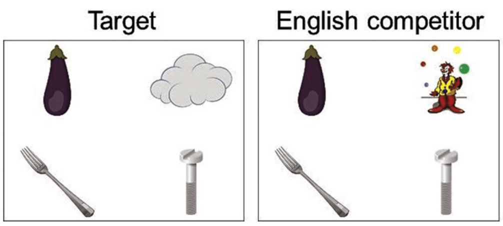


Figure 1.7: Example of a visual display. The figure shows two sample scenes used in the experiment: on the left, the target display includes the predictable object *cloud* among unrelated distractors; on the right, the English competitor display replaces the phonological competitor (*clown* and three distractors). *Note.* From Ito, A., Pickering, M. J., & Corley, M. (2018). Investigating the time-course of phonological prediction in native and non-native speakers of English: A visual world eye-tracking study. *Journal of Memory and Language, 98*, 1–11. © 2017 Elsevier. Reproduced with permission.

While Ito et al. (2018) provided evidence for a short-lived phonological prediction effect, the robustness of such effects was further questioned by Ito and Husband (2017). In their visual world eye-tracking study, Ito and Husband (2017) directly contrasted semantic and phonological prediction. Participants listened to highly constraining sentences such as “That dog looks so happy, wagging its ...” while viewing displays containing four objects: one

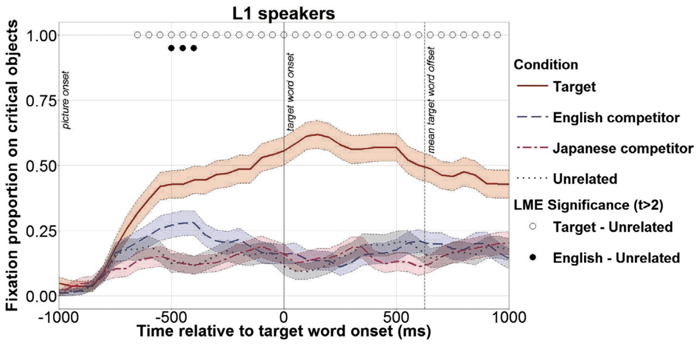


Figure 1.8: Time-course of fixation proportions. *Note.* From Ito, A., Pickering, M. J., & Corley, M. (2018). Investigating the time-course of phonological prediction in native and non-native speakers of English: A visual world eye-tracking study. *Journal of Memory and Language, 98*, 1–11. © 2017 Elsevier. Reproduced with permission.

critical object and three unrelated distractors. Depending on condition, the critical object was either the predictable target (*tail*), a semantic competitor (*paw*), a phonological competitor overlapping at onset (*table*), or an unrelated control item. The results showed clear anticipatory looks to the semantic competitor but no predictive looks to the phonological competitor. This finding suggests that while comprehenders routinely pre-activate semantic information related to the predicted target, phonological pre-activation is much less consistent. The authors suggest that this may arise because semantic prediction can operate at a more general level (e.g., predicting features related to *dog*), whereas phonological prediction necessarily requires committing to a specific word form. This process could be particularly demanding and perhaps less commonly used, especially given that highly predictable words are rare in natural language (Luke & Christianson, 2016).

Other studies have examined phonological prediction in languages with non-alphabetic scripts, where orthographic overlap is less likely to confound phonological effects. However, the findings are far from uniform. In Japanese, Ito (2019) used a printed-word visual-world variant to test whether listeners pre-activate orthographic and/or phonological form during spoken language comprehension. The author conducted two visual world experiments with native Japanese speakers in Tokyo, manipulating the script (kanji vs. hiragana) in which critical words were displayed.

Participants (N = 40 per experiment) listened to highly predictable sentences (mean cloze probability, 87%) while viewing displays of four printed words. In Experiment 1, words were presented in kanji (the logographic script), while in Experiment 2, the same words appeared in hiragana (a syllabic, phonologically transparent script). In both experiments, the orthographic competitor was orthographically similar to the target in kanji but phonologically dissimilar (e.g., 魚 /sakana/ 'fish' vs. 角 /tuno/ 'horn'), whereas the phonological competitor shared one mora with the target but was orthographically dissimilar in kanji (e.g., 魚 /sakana/ vs. 桜 /sakura/ 'cherry blossom'). Words appeared on the screen 1000 ms before the target word onset. Results showed a clear orthographic competitor effect in Experiment 1 (when words were presented in kanji), but no phonological competitor effect. In Experiment 2 (hiragana presentation), the pattern reversed: a phonological competitor effect emerged (e.g., さかな /sakana/ 'fish' vs. さくら /sakura/ 'cherry blossom'), but the orthographic competitor effect disappeared when the competitor was no longer visually similar to the target (e.g., さかな vs. つの /tuno/ 'horn'). Note that in hiragana, the phonological competitor was now also orthographically similar to the target by sharing the first character. Ito (2019) interpreted these findings as evidence that orthographic prediction depends on the availability of relevant orthographic information in the visual context. Moreover, this orthographic activation could be driven by bottom-up priming from the displayed words, rather than a purely predictive process, or participants could even have mistaken the orthographic competitors for the target.

To test whether such effects generalize from printed words to pictures, Ito and Sakai (2021) extended this work by replacing printed words with pictures of objects, thus removing any direct orthographic input from the visual display itself. Using the same sentence materials as Ito (2019), they tested 57 native Japanese speakers in Tokyo (Experiment 1) and 56 in Berlin (Experiment 2). The latter group had reduced daily exposure to Japanese (43% of the time, compared to nearly 100% for the Tokyo group). Remarkably, even without printed words, the Tokyo group still showed anticipatory fixations to orthographic competitor objects, suggesting that listeners mentally activated the written form of the predicted word and this activation influenced their eye movements toward objects whose names were orthographically related in kanji. However, neither group showed evidence of phonological competitor effects. This seems inconsistent with Experiment 2 from Ito (2019), which found phonological competitor preactivation with the hiragana script, though it is important to note that those phonological competitors were also orthographically related in the first character, which could have contributed to the effect. The authors also attribute this discrepancy in part to weak phonological relatedness (one-mora overlap may be insufficient to drive eye movements in a visual world paradigm with pictures). The Berlin group showed weaker overall prediction effects, with orthographic competitor effects emerging only for items with very high orthographic similarity. These results suggest that frequent language exposure strengthens lexical prediction but does not necessarily extend to phonological levels. Thus, the Ito and Sakai (2021) study, contrary to the initial interpretation of Ito (2019), found that for orthographic form preactivation, having orthographic information present in the visual display is not a prerequisite. Furthermore, the orthographic activation in Experiment 1 of Ito (2019) could not have been due to simple bottom-up priming from the presented visual context or to the orthographic competitor being mistaken for the target.

In contrast to the Japanese findings, Li et al. (2022) found phonological prediction effects in Mandarin Chinese. In Experiment 1, 42 native Mandarin speakers (Beijing university students) listened to highly predictable sentences (mean cloze 92.3%) while viewing displays of four pictures. The phonological competitor shared the initial syllable and tone with the target word (e.g., /shu1bao1/ 'schoolbag' vs. /shu1zi5/ 'comb'). The picture preview time was 2000 ms. Results showed clear anticipatory fixations to both semantic and phonological competitors. The phonological competitor effect emerged as early as 1400 ms before target onset, considerably earlier than effects reported in English studies (e.g., 300–500 ms before target onset in Ito et al., 2018).

This discrepancy was explained by the results of post-hoc cloze testing, which showed that the target words were already moderately predictable (33% cloze) 1400 ms before their actual onset. This suggests that the strong contextual constraints, coupled with the extended preview time (2000 ms compared to 1000 ms in Ito et al., 2018), allowed for very early predictions. Experiment 2 replicated the phonological effect with a shorter preview time (1000 ms) and different stimuli in which phonological competitors differed in tone from the target. Phonological effects still emerged, but later (400 ms before target onset), confirming that preview duration influences the time course but not the presence of phonological prediction.

Li and Qu (2023) replicated Li et al. (2022) (Experiment 1) while additionally measuring participants' verbal working memory capacity (WMC) using a reading span task. Forty participants were divided into high- and low-span groups. Results replicated the phonological and semantic competitor effects in both high- and low-WMC groups. Interestingly, WMC modulated the time course of semantic prediction (earlier onset in high-WMC participants) but did not affect phonological prediction onset timing (only the duration of the effect was slightly longer for the high-span group). Importantly, both groups showed prediction. Thus, low WMC does not cancel out prediction but only modulates its time course, which is consistent with the ERP results by Otten et al. (2007) reported earlier, who also found that WMC did not determine whether grammatical gender predictions were generated, but rather influenced how prediction-inconsistent information was subsequently processed (with low-span readers showing an additional late negativity reflecting greater difficulty integrating unpredicted words).

The contrast between these findings – clear phonological effects in Mandarin, but not in Japanese – suggests that phonological prediction is highly sensitive to multiple interacting factors. First, the strength of phonological overlap matters: Mandarin studies used full syllable and tone overlap, whereas Japanese studies used minimal one-mora overlap. Single-mora overlap may simply be insufficient to drive eye movements in the visual world paradigm, even if phonological form is pre-activated to some degree (Ito and Sakai, 2021). Second, the characteristics of the writing systems differ in ways that may influence processing: Japanese kanji characters often have multiple possible readings, potentially creating competition or uncertainty that interferes with phonological prediction (Ito, 2019). Third, preview time appears to influence the time course of effects: the 2000 ms preview in Li et al. (2022) (Experiment 1) allowed phonological effects to emerge much earlier than the 1000 ms preview used in Experiment 2. Fourth, the use of printed words versus pictures may influence competitor effects. Fifth, the strength of contextual constraint in the materials could influence predictive processing, with 92.3% cloze probability in the Chinese studies compared to 87% in Japanese studies. Finally, individual differences in language exposure and cognitive resources also modulate prediction. Ito and Sakai (2021) found that reduced daily exposure to Japanese (in the Berlin group) weakened both target and orthographic competitor effects. Li and Qu (2023) found that verbal WMC influenced mainly semantic prediction, suggesting that predictive processing draws on cognitive resources and that different representational levels (semantic and phonological) may utilize cognitive resources differently. However, none of these studies systematically examined participants' broader linguistic repertoires (e.g., bilingualism, literacy in other scripts, or second-language proficiency), which may further modulate predictive processing.

In sum, the available evidence from non-alphabetic languages indicates that phonological prediction can occur under supportive conditions, namely strong phonological overlap, sufficient processing time, and highly predictable contexts, but is not consistently observed across all paradigms. The observed variability reflects the interaction of linguistic features (e.g., phonological overlap strength, script-specific characteristics), methodological parameters (timing, display format, contextual constraint), and individual differences (language exposure, cognitive resources, potentially linguistic repertoires). Consistent with this interpretation, Ito (2024) conducted a meta-analysis of 20 visual-world eye-tracking experiments and found a small but reliable predictive phonological competitor effect that emerged briefly before the target onset. This effect corresponded to about a 5% increase in looks to phonologically related items compared to unrelated distractors and lasted roughly 600 ms. The size and timing of this effect were modulated by several experimental factors. Stronger effects occurred in highly constraining sentences (cloze probability above 90%) and when the target and competitor shared greater phonological overlap (e.g., several shared phonemes or an entire syllable in tonal languages). Differences in preview duration (1000 vs. 2000 ms) mainly affected when, rather than how strongly, the effect appeared: longer previews tended to shift the effect earlier in time, but the overall magnitude and duration remained comparable across studies. Interestingly, the predictive phonological competitor effect was slightly larger for picture than for printed-word stimuli, contrary to the pattern typically observed in spoken-word recognition (Huettig & McQueen, 2007), suggesting that the printed-word paradigm is not necessarily more sensitive to phonological prediction. Overall, these results indicate that while phonological pre-activation is possible, it is weaker, shorter-lived, and more dependent on contextual and design factors than semantic prediction, surfacing only under optimally supportive conditions.

Production evidence (context-driven picture naming). Although most of this literature focuses on comprehension, production paradigms can offer evidence about whether predictions reach phonological levels. The sentence-to-picture naming task is one example. In a series of three experiments, Drake and Corley (2014) presented participants with auditory sentence fragments with highly predictable continuations (e.g., "He managed to fix the drip from the old leaky ...") and asked them to name a picture presented immediately at the offset of the audio. The relationship between the sentence-predicted completion and the picture's name was manipulated so that the picture name either corresponded to the predictable word (tap–TAP), overlapped phonologically either at onset (tap–TAN) or at offset (tap–CAP), or had no overlap (tap–CONE). Across experiments, naming was facilitated only in the matching (exact-target) condition, whereas picture names that overlapped phonologically with the predicted word showed neither facilitation nor inhibition.

The authors argued that listener-generated predictions during comprehension do not activate speech-motor representations at the level required to produce phonological facilitation effects in overt naming. This finding might seem to contradict the neurocognitive evidence presented earlier showing speech-motor activation during comprehension (e.g., Scott et al. 2009; Watkins et al. 2003). However, Drake and Corley propose that speech-motor activation during comprehension may serve different functions. While the motor system may support prediction of when upcoming material will occur (the "how" pathway), their results suggest it does not necessarily encode the phonological content of predicted words in a production-ready format (the "what" pathway). Importantly, this account does not preclude phonological-form preactivation in comprehension (e.g., DeLong et al. 2005). Phonological expectations could exist at a representational level accessible to comprehension processes without necessarily being formatted for articulatory output. The production results thus provide a boundary condition: exact-match facilitation can occur even when partial phonological overlap alone produces no detectable effect. Consequently, null results for phonological competitor activation should be treated cautiously, as they do not necessarily imply that phonological preactivation is absent; rather, predictions generated during comprehension may not automatically transfer to the production system.

In this section, I examined prediction within typical adult comprehension, focusing on how anticipatory processes operate across semantic, grammatical, and phonological levels. The evidence suggests that prediction operates in a graded manner, with its strength and timing varying across levels of linguistic representation. While semantic and morphosyntactic features are routinely pre-activated, phonological prediction is weaker, shorter-lived, and more dependent on contextual and methodological factors. The following section explores how these mechanisms manifest across different populations, including second-language speakers and adults with developmental dyslexia.

1.4 Spoken Language Comprehension in Different Populations

Building on the motivations outlined in the General Overview, this section examines prediction in populations whose linguistic experience and processing profiles differ from those of typical L1 adults. I first review findings on prediction in second-language (L2) speakers and then consider evidence from individuals with developmental dyslexia, focusing on how semantic, morphosyntactic, and phonological information is anticipated in these groups.

1.4.1 Second-Language (L2) Speakers

By "L2 speakers," I refer to individuals who acquired a given language after their first language, usually under circumstances different from those of the first language (i.e., not through caregiving in the home), and thus distinct from simultaneous bilingual exposure in early childhood (Schlenter, 2023).

The terminology used to refer to individuals who know multiple languages has been subject to considerable debate in applied linguistics and multilingualism research. In this dissertation, the label "L2" does not imply that the language in question is literally a speaker's second language chronologically; rather, it encompasses any additional language(s) learned subsequent to the first. This broad usage reflects the prevailing practice in psycholinguistic research, where "L2" is typically employed as a general shorthand for non-native language processing,[5] even though most studies

[5]The term "non-native" is used here in a conventional sense to refer to speakers who acquired the target language after their first language rather than simultaneously. This operational use does not assume a strict native–non-native dichotomy, which is widely recognised as socially constructed rather than categorical (V. Cook, 1999; Davies, 2003). Throughout this dissertation, the terms "native" and "non-native" are used strictly in an acquisition-based sense, referring to speakers' first-acquired (L1) and subsequently acquired (L2) languages, without evaluative or proficiency-related implications.

do not explicitly define the term or clarify the assumptions underlying their use of it (e.g., Hopp, 2013; Kaan, 2014; Lew-Williams & Fernald, 2010). Hammarberg (2010) identifies two distinct terminological models in multilingualism research. The "linear model" uses ordinal counting (L1, L2, L3, L4...) to classify languages according to their chronological order of acquisition whereas the "two-level model" makes a broader distinction between the native language(s) (L1), established at a certain level during infancy, and all subsequently learned languages (L2). Many scholars investigating third language (L3) acquisition argue that treating all non-native languages under the umbrella term "L2" risks obscuring important differences between learning a second versus a third or additional language (Bui, 2023; Cenoz et al., 2001; De Angelis, 2007). From this perspective, L3 learning is qualitatively distinct because it involves unique cognitive and motivational dynamics not present in L2 learning (Bui, 2023; Hammarberg, 2010).

Despite the theoretical merit of these distinctions, I adopt the two-level model and use "L2" as a broad term for several reasons. First, the focus of the present research is on comparing native versus non-native language processing rather than on the chronological order of acquisition. Second, as Hammarberg (2010) acknowledges, linear ordering of languages becomes difficult to apply in many real-world contexts. This is particularly relevant for the present study: many participants in South Tyrol first acquire Tyrolean at home, then develop literacy in Standard German at school, and subsequently learn Italian (see Section 1.5). Whether Italian should therefore be classified as their L2 or L3 is not straightforward. Similar complexities arise for participants with other language backgrounds in the English experiment. For these reasons, I use "L2" as a broad category that includes all languages acquired after one's native language(s).

Semantic prediction in L2. A growing body of work has investigated whether L2 speakers, like L1 comprehenders, use contextual cues to anticipate upcoming words during spoken language comprehension. As discussed in the previous section, it is useful to distinguish between target (or contextual) prediction (i.e., the anticipation of a specific high-cloze target word) and semantic prediction, defined as the spread of activation to its semantic competitors. Most studies on L2 prediction have concentrated on the former, showing that second language speakers can anticipate targets from constraining contexts, though the timing and strength of effects often vary across tasks, linguistic cues, and language profiles.

Chambers and Cooke (2009) examined verb-driven anticipation in English L1, French L2 speakers with varying proficiency in French. The study employed interlingual near-homophones (e.g., French *poule* 'chicken' and English *pool*) to measure cross-language lexical competition during spoken sentence comprehension. Listeners' eye movements revealed temporary looks on the English competitor word (e.g., *pool*) when hearing the French target noun (*poule*), demonstrating cross-language lexical activation. However, competitor fixations were dramatically reduced when prior sentence context was incompatible with the competitor (e.g., *Marie va nourrir...* 'Marie will feed...' ruled out the pool). Following such constraining verbs, listeners fixated the only feedable object (e.g., the chicken) significantly more than after neutral verbs like *décrire* 'describe', demonstrating successful verb-based prediction in L2. Importantly, while participants' French proficiency reliably predicted their ability to use predicate information to anticipate target referents (accounting for 38% of the variance, $p < .01$) and their target selection accuracy, it had no significant effect on the extent of interlingual competition from their L1 English. This dissociation suggests that increased proficiency in the L2 facilitates the speed and efficiency with which contextual constraints are applied during real-time comprehension, but does not independently reduce interference from the L1. However, this study tested only L2 comprehenders and did not provide a direct comparison with L1 processing.

By contrast, Dijkgraaf et al. (2017) provided a more comprehensive comparison by testing Dutch–English bilinguals in both their L1 Dutch and their L2 English, alongside monolingual English controls. The bilingual participants (mean age 24 years) were dominant in Dutch, with relatively high proficiency in English, though they reported using Dutch approximately 73% of the time versus only 17% for English and had no immersion experience in English-speaking environments. Monolingual English controls (mean age 20 years) were matched to the bilinguals on L1 proficiency based on LexTALE scores (Lemhöfer & Broersma, 2012). Participants heard sentences with constraining verbs, such as “Mary knits a scarf,” versus neutral verbs like “Mary loses a scarf,” while viewing displays containing the target (scarf) and three unrelated distractors (all objects could be lost, but only the scarf was knittable). Bilinguals showed significantly more fixations toward the target object in the predictive condition before target word onset, both in their L1 Dutch and L2 English. Moreover, the magnitude of this anticipatory effect was statistically indistinguishable between the bilinguals’ L1 and L2 performance, and neither differed from monolingual controls.

Interestingly, unlike Chambers and Cooke (2009), English proficiency (LexTALE scores) did not modulate the strength of prediction effects in Dijkgraaf et al.’s study. This discrepancy may reflect differences in both the proficiency ranges sampled, as well as how proficiency was operationalized. While Chambers and Cooke used a composite proficiency measure that weighted self-rated listening (35%), speaking (10%), reading (20%), and writing (5%) abilities, along with years of immersion experience (20%) and current L2 exposure (10%), thereby capturing multiple dimensions of language experience relevant to real-time spoken comprehension, Dijkgraaf et al. relied solely on LexTALE scores, which assess only receptive vocabulary knowledge. Additionally, Dijkgraaf et al. report that their participants were relatively homogeneous in their L2 proficiency (mean LexTALE: 78.5%, SD = 10.49), potentially creating a ceiling effect.

The question of whether and how proficiency modulates L2 prediction was also addressed by Tagliani et al. (2025), who tested Italian learners of English at three proficiency levels (B1, B2, and C1/C2) alongside native English controls using a visual world paradigm. Participants heard sentences such as “A girl is wearing a dress” while viewing displays containing a target object (e.g., *dress*) and three distractors. The task manipulated both semantic cues (only one object was semantically plausible given the verb, e.g., only the dress was wearable) and phonological cues (the target could rhyme with the sentence subject, e.g., "A parrot is eating a carrot"). All proficiency groups showed anticipatory fixations to semantically appropriate targets during the verb window, confirming verb-based prediction in L2. However, low-intermediate (B1) English learners were slower to fixate the target compared to more advanced learners and native speakers. Moreover, when examining the integration of multiple cues (semantic + phonological), lower-proficiency learners (B1 and B2) were slower to identify targets when both cues were present compared to semantic cues alone, whereas advanced (C1/C2) learners showed native-like integration patterns. The authors attribute this difficulty to increased cognitive demands of cue integration in learners with more limited L2 processing resources. Moreover, Borovsky et al. (2012) demonstrated that vocabulary knowledge, independent of age, predicted the speed and efficiency of anticipatory processing in both children (ages 3-10) and adults during spoken sentence comprehension.

At the same time, other studies report modest L2 slowdowns. Ito et al. (2018), for example, found that both native speakers and Japanese learners of English generated anticipatory looks to predictable objects, but the onset of prediction was slightly later in the L2 group. Moreover, L2 prediction was found to be strongly shaped by prior linguistic experience. Van Bergen and Flecken (2017) examined whether L2 learners use Dutch placement verbs, which specify object position (*zetten* 'put.STAND' vs. *leggen* 'put.LIE'). Advanced German L1–Dutch L2 speakers, whose L1 also encodes object position information in placement verbs, showed immediate anticipatory fixations comparable to native Dutch listeners. In contrast, English and French L1–Dutch L2 speakers, whose L1 does not mark this contrast (the equivalent verbs *put* or *mettre* leave object position underspecified), failed to anticipate the object despite demonstrating explicit knowledge of Dutch placement verbs. These results indicate that predictive use of verb semantics in the L2 depends on cross-linguistic similarity: when the relevant semantic contrast exists in the L1 (as in German), it facilitates prediction, but when it does not (as in English or French), L1 transfer constrains anticipatory processing.

Moreover, when prediction requires integrating multiple cues, L1–L2 differences become even more apparent. As already noted in Tagliani et al. (2025), lower-proficiency L2 learners showed significant costs when integrating phonological and semantic information, whereas advanced learners performed similarly to native speakers. This pattern of increased difficulty with multiple cue integration extends beyond phonological-semantic combinations to other types of information integration. Corps et al. (2023) combined verb-based associations with speaker gender stereotypes. Participants listened to sentences such as "I would like to wear the nice…" produced by male or female speakers, while viewing stereotypically masculine (e.g., *tie, drill*) and feminine (e.g., *dress, hairdryer*) objects. Both L1 and highly proficient L2 participants anticipated objects consistent with the verb semantics (e.g., *tie* or *dress* after *wear*) at a similar time course. But consistent predictions that also integrated speaker gender (e.g., preferring *tie* for a male speaker) emerged later in the L2 group. The authors argue that such perspective-based predictions are less automatic in a second language, as they depend on resources that may be more limited for non-native speakers - whether due to reduced efficiency in self–other adjustments (Peters et al., 2015), greater inhibitory demands, or slower access to stereotypical knowledge. The authors verified that participants could correctly match each voice to its corresponding speaker before the experiment, suggesting that delays in perspective-based prediction were not due to difficulty identifying the speaker's gender.

Direct evidence for semantic competitor pre-activation in L2 is very limited. Using the visual world paradigm, Dijkgraaf et al. (2019) investigated whether the strength and time course of semantic prediction differ between L1 and L2 by testing the same 50 Dutch–English bilinguals dominant in Dutch (mean age 19 years, highly proficient in both languages) in both of their languages. Participants listened to audio sentences varying in cloze probability (mean cloze: 0.71 in Dutch, 0.68 in English) while viewing displays containing either a target picture or a semantic competitor, whose semantic distance from the target was quantified using word co-occurrence statistics in text corpora (e.g., bottle–glass = 0.41 vs. bottle–fridge = 0.56), along with three unrelated objects. Pictures appeared 500 ms before target onset to minimize visual priming effects. Participants were instructed to look anywhere they wanted as long as their eyes remained on the screen. Participants showed anticipatory fixations not only to predictable targets but also to their semantic competitors before target onset in both their L1 (Dutch) and L2 (English). Time-course analyses revealed that fixations to the target became significant approximately 250 ms before target onset, while increased looks to semantic competitors emerged around 100 ms before target onset in both languages. However, despite similar temporal dynamics, the magnitude of the prediction effect was stronger in L1 than in L2. Moreover, competitor fixations varied with semantic distance (participants looked more to closely related competitors), and this effect was significantly larger in L1 than in L2 throughout the prediction window. The authors interpret these findings as reflecting weaker lexico-semantic mappings and slower spread of activation in the L2, rather than a qualitative absence of semantic prediction. This pattern supports the weaker links hypothesis, which posits that L1 words have stronger and more densely interconnected representations in semantic memory than L2 words (Gollan et al., 2008).

Converging evidence for semantic prediction differences comes from an ERP study by Ito et al. (2017a), who examined whether L2 speakers pre-activate upcoming meaning during reading. Twenty-four Spanish L1, English L2 speakers read English sentences word by word at two presentation rates (stimulus-onset asynchrony, SOA = 500 ms and 700 ms). Each high-cloze sentence (e.g., "The student is going to the library to borrow a...") was followed by either the expected word (*book*), a semantically related word (*page*), a form-related word (*hook*), or an unrelated word (*sofa*). Semantically related words elicited smaller N400 amplitudes than unrelated words at both presentation rates, suggesting facilitated semantic processing. However, this

N400 reduction did not depend on the cloze probability of the expected word unlike in native speakers tested with the same materials in a previous study by the authors (Ito et al., 2016, see also Section 1.3.1). Because a genuine prediction effect should scale with predictability, the authors concluded that the observed N400 reduction reflected easier integration of plausible words rather than pre-activation of semantic features. These findings suggest that L2 readers were not pre-activating semantic information, but rather benefited from semantic relatedness during integration.

Findings such as these align with Schlenter's (2023) review of more than forty studies: L2 speakers successfully predict upcoming linguistic input based on lexical semantics, but effects are typically delayed or attenuated, particularly when prediction requires integrating multiple cues beyond simple verb–object associations. Bovolenta and Marsden (2022) similarly emphasize that variability between L1 and L2 speakers (and even within each group) increases as prediction becomes more complex, whereas sensitivity to simple contextual predictability is the most robust and least divergent across groups. Thus, while L2 comprehenders clearly use context to anticipate likely referents, the extent and timing of semantic competitor activation are less robust than in L1 comprehension. Moreover, the extent to which their predictions spread to semantically related alternatives remains understudied.

Morphosyntactic prediction in L2. Beyond semantics, an important question is whether L2 comprehenders, like native speakers, can use morphosyntactic cues such as grammatical gender, number, or case to anticipate upcoming words. Evidence suggests that L2 speakers can use these cues predictively, but effects are typically weaker, delayed, or more variable than in L1 comprehension, and are strongly modulated by lexical knowledge, proficiency, and cross-linguistic congruency.

One of the earliest demonstrations comes from Hopp (2013), who examined whether advanced L1 English learners of German could use determiner gender marking to predict upcoming nouns in a visual-world eye-tracking task. Native German speakers showed anticipatory looks toward the gender-congruent referent upon hearing the determiner (e.g., *der*/*die*), before the noun was spoken. L2 learners also displayed anticipatory effects, despite their L1 lacking grammatical gender. However, the strength of their predictions correlated with their lexical gender knowledge, measured in a separate production task: learners who were more accurate in gender assignment also showed stronger anticipatory eye movements. These findings suggest that L2 speakers' predictive processing depends on the stability and accessibility of lexical gender nodes in the bilingual lexicon.

A complementary perspective comes from Lew-Williams and Fernald (2010), who directly compared the online use of gender cues by adult L1 English–L2 Spanish learners (with an average of five years of classroom instruction) with L1 Spanish adults and children acquiring Spanish as their first language. In this paradigm, participants viewed two pictures with names that either shared the same grammatical gender (e.g., *la pelota* 'ball', *la galleta* 'cookie') or differed in gender (e.g., *la pelota*, *el zapato* 'shoe'), while hearing sentences such as *Encuentra la pelota* ('Find the ball') or *¿Dónde está la pelota?* ('Where is the ball?'). Because participants could not know in advance which picture would be named, their initial gaze was equally likely to fall on either image. The critical measure was the latency with which they shifted their gaze from the distracter to the target picture once the article–noun sequence began. Faster shifts on different-gender trials, where the article uniquely identified the noun, would indicate that listeners used gender marking incrementally to constrain lexical access. Although this does not constitute anticipatory prediction strictly speaking (since the noun onset was already available), it reflects rapid, online use of grammatical gender information during spoken language comprehension. Native Spanish adults and children showed precisely this pattern, orienting to the target more rapidly when the article provided an informative, prenominal cue to the noun's gender. In contrast, L1 English learners of Spanish showed no facilitation on different-gender trials, even though they had explicit knowledge of gender agreement rules. In two further experiments using novel nouns, the authors manipulated whether article–noun pairings were consistent across training and testing. When exposure involved novel nouns and the same article–noun pairings in training and test (e.g., participants learned *la catela* (feminine object) and were later tested with *la catela*), L2 learners showed a small facilitation effect, suggesting that repeated fixed pairings can support use of the cue. However, when participants were trained on indefinite articles (*una/un*) and later tested on definite articles (*la/el*) with the same nonce nouns (e.g., *la catela* after *una catela* training), L1 adults generalized and used the definite article incrementally, whereas L2 adults did not. This finding suggested that L2 learners could rely on associative links between specific article–noun sequences, but not on the generalization of gender categories (e.g., the feminine marking 'la' predicts nouns from the same class). Lew-Williams and Fernald argued that adult L2 learners often possess explicit (declarative) knowledge of gender rules but lack automatic, proceduralized links between determiners and nouns that support rapid incremental processing. Classroom instruction emphasizes rule learning and limited practice, which does not yield strong article–noun associations that native speakers accrue from early and varied input. In addition, the authors argued that cross-linguistic transfer from a first language without grammatical gender further limits L2 speakers' ability to deploy gender marking in real time.

Building on this, Hopp and Lemmerth (2018) tested Russian adult learners of German, two languages with three gender categories but with differences in how gender is realized: Russian encodes gender through postnominal suffixes, whereas German encodes it primarily prenominally on determiners, although some noun-internal derivational suffixes (e.g., *-heit, -ung*) are strongly associated with a particular gender. Using eye-tracking, they found that advanced learners showed native-like predictive use of gender across conditions. High-intermediate learners, however, displayed asymmetries. In adjective contexts, where Russian and German are syntactically congruent (both mark gender on suffixes), learners predicted nouns regardless of whether the noun carried the same gender in Russian and German (i.e., irrespective of lexical congruency). In article contexts, which are syntactically incongruent across the two languages, predictive use of gender occurred only when lexical gender overlapped between Russian and German. Thus, cross-linguistic similarity at both the lexical and syntactic level strongly determined whether L2 comprehenders could exploit gender cues predictively.

Proficiency effects on prediction have been further examined by Dussias et al. (2013), who investigated whether second-language learners exploit grammatical gender information during comprehension in a manner comparable to native speakers, using the visual-world eye-tracking paradigm. They tested three participant groups: sixteen Spanish monolinguals, eighteen English-speaking learners of Spanish (divided into high- and low-proficiency subgroups based on standardized Spanish proficiency test scores), and sixteen Italian learners of Spanish. Both English and Italian participants were late L2 learners who had acquired Spanish as adults. All groups demonstrated a high level of explicit knowledge of gender agreement in offline production and comprehension tasks. During eye tracking, participants viewed two pictures on the screen whose names were either of the same or different grammatical gender (e.g., *la pera* 'the pear' and *el pozo* 'the well') while listening to sentences in which one picture was named. Sentences were syntactically and semantically varied, with the target noun appearing either in the middle or at the end of the sentence (e.g., the target *el reloj* 'the clock' in *El estudiante estaba dibujando el reloj que vio ayer* 'The student was drawing the clock that he saw yesterday'). The critical measure was the latency (change-point) at which participants' eye movements converged on the target following the onset of the gender-marked article. Although the authors consider the results to be anticipatory, this shift usually coincided with the initial stages of noun recognition. Therefore, the observed effects can be more accurately described as a facilitatory use of gender marking rather than a strictly predictive one. Spanish monolinguals reliably used gender-marked articles to orient their gaze toward the correct referent as soon as gender information became available, indicating that gender served as a facilitative cue for lexical access. High-proficiency English–Spanish learners showed a qualitatively similar pattern, shifting gaze to the target earlier on different-gender than on same-gender trials for both masculine and feminine nouns. In contrast, low-proficiency English–Spanish learners showed no such facilitation, indicating that they had not yet integrated grammatical gender cues into their online language processing routines. Italian–Spanish bilinguals, whose L1 also encodes grammatical gender, displayed facilitative effects only for feminine nouns. The absence of facilitation for masculine nouns was attributed to the presence of a higher proportion of opaque masculine nouns in the materials (i.e., words without canonical masculine gender endings such as *-o*), which likely increased processing demands, and to participants' relatively low L2 proficiency. The authors interpreted these results as evidence that both L2 proficiency and cross-linguistic similarity determine the extent to which learners can exploit morphosyntactic cues during comprehension.

Comparable findings have been reported in child bilinguals. Bosch and Foppolo (2022) tested Italian–German bilingual children aged six to ten. Using a visual-world eye-tracking paradigm, they showed that children anticipated upcoming nouns based on gender-marked articles in both languages. However, when the same noun had incongruent gender values across the two languages (e.g., feminine in Italian but masculine in German), prediction was delayed – a "gender congruency effect." Moreover, predictive use of gender was modulated by language dominance, with faster prediction in the dominant language. These results highlight that even in simultaneous bilinguals, predictive morphosyntactic processing is sensitive to cross-linguistic influence and linguistic proficiency.

Other morphosyntactic domains appear to be even more challenging. Mitsugi (2017) examined whether English-speaking learners of Japanese could use case markers to anticipate upcoming verb structure in a verb-final language. Native Japanese speakers used nominative and dative markers to predict passive verbs before they appeared. L2 learners, however, did not anticipate structure from case markers and instead relied on verb morphology once the verb was encountered. Similarly, Hopp (2015) showed that German learners of English had difficulty using case marking to resolve temporary object–subject ambiguities during reading.

Taken together, these findings indicate that while L2 comprehenders are capable of morphosyntactic prediction, such effects are weaker compared to contextual and semantic prediction. This gives evidence that predictive use of grammatical cues depends on stable lexical–syntactic mappings, cross-linguistic overlap, and overall language proficiency. Recent reviews by Schlenter (2023) and Ito and Pickering (2021) agree on the conclusion that morphosyntactic prediction appears to be especially fragile in L2 comprehension: L2 speakers may rely more readily on semantic cues, but are less consistent in exploiting morphosyntactic information, especially when it is absent or differently realized in their L1. Ito and Pickering (2021) further argue that one possible explanation lies within the prediction-by-production framework (see Section 1.2.3). Since prediction is assumed to rely on production mechanisms that are less automatized in the L2, the predictive use of morphosyntactic cues in comprehension is particularly difficult. However, this raises an important question as to why such differences are less apparent for semantic prediction. One plausible explanation is that, unlike grammatical prediction, which depends on precise morphosyntactic representations, semantic prediction relies more on conceptual and associative networks that are largely shared across languages. Evidence from bilingual semantic memory research supports this view: studies using repetition priming (showing faster or more accurate responses to repeated concepts across languages) and false memory paradigms (where semantically related words that were not actually presented are falsely recalled or recognized across languages) consistently show that translation equivalents access common core-meaning representations across languages (Francis, 2020), and that semantic associations, such as relationships between categories and their members (e.g., fruit–strawberry) or between objects and their typical actions (e.g., scissors–cut), are shared across bilingual's two languages. In contrast, morphosyntactic structures are inherently language-specific, requiring the activation of L2-specific grammatical encodings that may not be as readily accessible or automatized, particularly when they differ from L1 structures.

Phonological prediction in L2. Finally, another question is whether L2 comprehenders can pre-activate the phonological form of an expected word. Phonological prediction has been widely debated even in L1 research, with mixed evidence as to whether comprehenders routinely anticipate form-based information or only under specific conditions (see Section 1.3.3).

In native speakers, DeLong et al. (2005) reported graded N400 effects on the English indefinite article *a/an*, showing that readers predicted whether the next word would begin with a consonant or vowel. However, subsequent replication attempts have yielded inconsistent results (see Section 1.3.3; Nieuwland et al., 2018). Briefly, the large-scale multi-laboratory replication by Nieuwland et al. (2018), which reproduced the materials and procedures of DeLong et al. (2005) with a substantially larger native sample, confirmed graded N400 effects on the nouns but not on the preceding articles, suggesting no evidence for article-level phonological prediction. Extending this paradigm, Martin et al. (2013) addressed whether L2 comprehenders also anticipate phonological form. English monolinguals and Spanish–English late bilinguals read semantically constraining sentences such as "Since it is raining, it is better to go out with *an umbrella* [expected] / *a raincoat* [unexpected]." For native readers, ERP results showed sensitivity to article congruency, replicating DeLong's findings. In contrast, L2 readers showed no anticipatory effect at the article, with N400 modulations only emerging at the noun itself. The authors concluded that while L2 comprehenders engage in lexical prediction, they do not reach a similar level of phonological form pre-activation as native speakers.

In contrast to the reading-based ERP findings, a study using the visual-world paradigm by Connell et al. (2021) tested whether L2 speakers can use the English indefinite-article alternation to anticipate an upcoming noun during spoken language comprehension. In their design, the sentence context did not predict a specific noun (e.g., "The boy found..."); and the article *a/an* was the only cue to the noun's initial phoneme. Both L1 and

L2 listeners made anticipatory fixations to the picture consistent with the article before noun onset, and the effect was modulated by proficiency: higher-proficiency L2 speakers showed earlier and stronger anticipatory looks. These results indicate that, in spoken language comprehension, L2 speakers can use the *a/an* alternation as a phonological cue to anticipate the next word.

While this shows that L2 listeners can anticipate from phonological information explicitly provided in the input, a more stringent test is whether L2 speakers also pre-activate the phonological form of a highly predictable word strongly enough for its competitors to be activated. As discussed in Section 1.3.3, Ito et al. (2018) tested both native and non-native speakers using a visual-world paradigm. Focusing here on the L2 results, Japanese learners of English heard constraining sentences such as "The tourists expected rain when the sun went behind the..." while viewing displays that included the target picture (*cloud*) alongside a phonological competitor (*clown*), with an average overlap of 2.9 phonemes. Native participants launched anticipatory looks to both the target and the phonological competitor, indicating pre-activation of form. By contrast, L2 listeners showed delayed anticipatory looks to the target and no evidence of competitor activation in the anticipatory window. Instead, they fixated phonological competitors only after target onset, a pattern consistent with priming rather than prediction, i.e., a bottom-up activation of phonologically similar words in response to the incoming speech signal. A second condition tested whether L2 speakers would activate phonological competitors from their L1 Japanese while processing English sentences. In this cross-linguistic condition, the name of the critical object in Japanese phonologically overlapped with the Japanese translation of the English target. For instance, when the English target was *cloud* (translated as *kumo* in Japanese) the corresponding competitor was *bear* (*kuma* in Japanese). Despite this overlap, participants showed no anticipatory fixations to the Japanese competitor. The authors propose several reasons for this pattern, including weaker lexical–phonological links in the L2 and slower lexical access that prevents prediction from reaching the form level. This account also aligns with prediction-by-production models (discussed in Section 1.2.3), according to which phonological form corresponds to the final stage of language production and thus the part most likely to be slowed down in L2 processing.

More recently, Yin and Pickering (2026) directly addressed the null findings from Ito et al.'s (2018) cross-linguistic condition by investigating whether the absence of L1 phonological prediction might reflect methodological limitations rather than an actual inability of L2 comprehenders to pre-activate L1 phonological forms. Yin and Pickering hypothesized that cross-linguistic phonological prediction might depend on language context – specifically, whether both languages are actively in use during the experiment. Using a visual-world eye-tracking paradigm, they tested whether highly proficient Mandarin-English bilinguals pre-activate L1 (Mandarin) phonological representations of predictable words while comprehending L2 (English) spoken sentences, manipulating language context across two experiments. Importantly, whereas Ito et al. used L1 Japanese competitors with only partial phonological overlap with Japanese translations of English targets (averaging 53% overlap), Yin and Pickering employed Mandarin homophones providing complete phonological overlap between competitors and target translations. For example, participants heard highly constraining English sentences such as "You should take an umbrella with you, because there will be heavy *rain* at three o'clock this afternoon" while viewing displays containing a target object (rain, *yu3* in Mandarin), a Mandarin homophone competitor whose name was phonologically identical to the target's Mandarin translation (feather, also *yu3* in Mandarin), and unrelated objects. The results revealed a striking context-dependency of cross-linguistic phonological prediction. In Experiment 1, in which participants were exposed exclusively to English throughout the session (a one-language context), there was no anticipatory fixation to the Mandarin homophone competitor before English target onset, indicating that prediction remained selective to the target language (English). In Experiment 2, however, where the session included both English and

Mandarin trials as well as instructions in Mandarin (a two-language context), participants showed significantly increased anticipatory fixations to the Mandarin homophone competitor before the onset of the English target. The authors conclude that L2 comprehenders can pre-activate L1 phonological representations of predictable words, but only when both languages are contextually relevant. They propose that in two-language contexts, uncertainty about which language will be encountered next leads to non-selective lexical prediction, with the comprehension system activating candidate words in both languages. In contrast, in one-language contexts, the certainty that only the target language will be used results in language-selective prediction. However, these findings also raise questions about the order of form preactivation. As Yin and Pickering note, their design did not include L2 phonological competitors, leaving open whether L2 comprehenders directly pre-activate L1 phonological representations of predictable words, whether they first pre-activate L2 forms which then activate L1 translation equivalents, or whether L1 forms are activated first and subsequently trigger L2 translations.

While Ito et al. (2018) and Yin and Pickering (2026) examined whether L2 comprehenders pre-activate the phonological form of a predictable target strongly enough for its competitors to be activated, more recent work has addressed a complementary question of whether L2 comprehenders can use phonological cues from the input itself to generate predictions. As discussed in the section on semantic prediction above (Section 1.4.1), Tagliani et al. (2025) investigated this issue by examining how Italian learners of English integrate semantic and phonological cues during spoken comprehension. Focusing on their phonological findings, participants heard sentences designed to provide either only semantic cues (e.g., “A girl is wearing a dress”, with displays containing a single wearable object), only rhyme-based phonological cues (e.g., “A king is wearing a ring”, with displays containing several wearable objects but only one rhyming with the target), or both cue types combined (e.g., “A parrot is eating a carrot”, with one wearable object among distractors that also rhymes). The authors used rhyme as a phonological manipulation because shared rime units (e.g., king–ring), on a par with cohort competitors sharing initial phonemes, were found to activate overlapping phonological representations, as shown in visual-world studies where rhyme competitors elicit increased fixations during spoken word recognition (e.g., Allopenna et al., 1998; Magnuson et al., 2003). Native English speakers showed strong anticipatory looks driven by both cue types, although rhyme effects were weaker than semantic ones. L2 learners, by contrast, relied primarily on semantic cues and on the combination of semantic+phonological cues, but were consistently less sensitive to rhyme-based information alone. Proficiency modulated these effects: higher-proficiency learners showed clearer rhyme-based anticipation than lower-proficiency learners, yet they still did not reach native-like levels. As the authors conclude, while the capacity to exploit phonological information gradually develops with increasing proficiency, it remains weaker, slower to emerge, and more dependent on available processing resources than predictive use of semantic information in L2 comprehension.

Taken together, these studies indicate that phonological cues tend to be less efficiently exploited in L2 comprehension. Whereas semantic and morphosyntactic information is often used predictively, pre-activation of phonological form is inconsistent, weaker, and strongly modulated by proficiency and language context. Within prediction-by-production accounts (e.g., Ito & Pickering, 2021; Pickering & Gambi, 2018), this pattern is explained by the assumption that phonological prediction depends on activating later stages of the production system. Because these stages involve more detailed form encoding, they are more resource-demanding and therefore particularly susceptible to slowdowns in a less automatized L2.

Cognates and cross-linguistic influence. A question that has been a matter of concern for many scholars is how the bilingual language system is organized: whether lexical access is language-selective (i.e., restricted to the target language) or non-selective, such that representations from both languages become co-activated even when processing takes place in a single language. Early accounts proposed some degree of selectivity, assuming that bilinguals could efficiently restrict activation to the relevant lexicon (e.g., Scarborough et al., 1984). However, a large body of experimental evidence now supports non-selective access, where cross-linguistic overlap at the orthographic, phonological, and semantic levels can influence processing across tasks and modalities (Dijkstra & Van Heuven, 2002; Duyck, 2005; Marian et al., 2003; Schwartz et al., 2007). At the semantic level in particular, research on bilingual memory has demonstrated that translation equivalents and related words access shared conceptual representations across languages (Francis, 2020), suggesting that the bilingual lexicon is integrated at the level of meaning even if surface-form representations remain partially distinct.

At the same time, non-selectivity is not absolute: several studies have demonstrated that cross-linguistic information can be strategically suppressed when it is unhelpful for the task at hand. For instance, Mercier et al. (2016) showed that bilinguals who had just spoken in a different language exhibited reduced cross-language competition in a subsequent comprehension task, consistent with global inhibition of the non-target system. This distinction between global and local inhibition has been formalized in theoretical accounts: global control refers to the suppression of an entire language system depending on context, whereas local control applies more narrowly to specific competing lexical items (De Groot & Christoffels, 2006). Converging evidence for such context-driven suppression comes from ERP studies showing long-lasting inhibition of the dominant L1 during L2 production (Misra et al., 2012) as well as from immersion research demonstrating reduced L1 access when learners are immersed in an L2 environment (B. J. Levy et al., 2007; Linck et al., 2009).

The scope of non-selectivity may vary across representational levels and across speakers. For example, as mentioned above, Ito et al. (2018) found that Japanese learners of English did not pre-activate Japanese words that overlapped phonologically with highly expected English targets, suggesting that cross-linguistic phonological pre-activation may not occur in L2 comprehension. Yet even in L1, evidence for phonological pre-activation during comprehension remains mixed (see Section 1.3.3), raising the possibility that form-level prediction is generally difficult to detect and that its absence does not necessarily contradict the broader non-selectivity account. Furthermore, the degree of cross-language activation appears to depend on language proficiency and experience. Branzi et al. (2021) investigated this question with Chinese native speakers who varied in their proficiency in L2 English. In a semantic relatedness judgment task conducted entirely in English, the authors found that cross-language activation of Chinese, contrary to expectations, increased with L2 proficiency. More proficient bilinguals showed a stronger character repetition effect, i.e., they were more influenced by phonological overlap in the Chinese translations of the English words, whereas lower-proficiency participants showed evidence of L1 inhibition. This suggests that, rather than inhibiting their first language to optimize L2 processing, highly proficient bilinguals activate phonological representations from both languages in parallel. According to the authors, this reflects a qualitative shift in control mechanisms with increasing proficiency: less proficient speakers rely on inhibition to minimize interference from L1, whereas more proficient bilinguals develop stronger bidirectional connections between their L2 lexicon and the shared conceptual–semantic system, leading to more efficient comprehension.

One of the most compelling demonstrations of such cross-linguistic influence comes from cognates – lexical equivalents that share both form and meaning in distinct languages (e.g., English *piano* – Italian *piano*; Payesteh and Pham, 2022). Cognates typically show faster recognition and production latencies than noncognates, a pattern known as the **cognate facilitation effect** (Costa et al., 2000). In a paired-associate learning paradigm with Dutch–English word pairs, in which participants learn L2 words by repeatedly pairing each with its L1 translation, De Groot and Keijzer (2000) demonstrated that cognates are learned faster, retrieved more accurately, and forgotten less rapidly than noncognates. Their experiments show that shared form–meaning structure facilitates the establishment of new L2 word–concept links. Although their study focused on vocabulary learning rather than online comprehension, their results suggest that cognates reduce learning load because they can be mapped onto pre-existing L1 representations, and this mapping advantage likely persists during real-time processing. A parallel pattern is observed in spoken production. In a picture-naming study, Hoshino and Kroll (2008) found that bilinguals named cognate pictures faster and more accurately than noncognates, even when their two languages used different scripts (e.g., Japanese–English). Because no written forms were presented, this facilitation must have arisen from cross-language phonological activation during speech planning rather than from visual orthographic overlap. However, the direction of this effect in production can vary depending on the degree and type of cross-linguistic overlap. In production tasks such as word naming, when the stimulus is a written word rather than a picture, cognates with high orthographic similarity but divergent phonological forms (e.g., English *base* /beɪs/ vs. Spanish *base* /ˈba.se/) can produce interference rather than facilitation (Jared & Kroll, 2001; Schwartz et al., 2007). This occurs because such cognates activate competing phonological codes from both languages, and inhibiting the competing alternative from the non-target language delays articulation, particularly for less proficient L2 speakers (Jared & Kroll, 2001).

Recent evidence from L2 vocabulary research further illustrates the strength and asymmetries of cognate facilitation. In a study on Italian learners of English (aged 15–19, proficiency levels A2–C1), Vender and Nardon (2023) examined how cognate status affects both orthographic and phonological lexical decision performance using an adapted V_Yes/No test. Participants were presented with both real English words (e.g., *elegance*) and pseudowords (e.g., **rudge*) in either written or auditory form, and were asked to indicate which items they recognized as real English words. The test included cognates (words with formal similarity to Italian equivalents, e.g., *remove* – Italian *rimuovere*) as well as non-cognates. The authors found that cognates were recognized more accurately across both modalities and all proficiency levels, but the facilitation was markedly stronger in the orthographic condition. Interestingly, formal similarity also led to higher rates of false alarms: pseudowords resembling Italian words (e.g., **intimant*) were more often incorrectly accepted as real English words, particularly in written form (Vender & Nardon, 2023). These findings indicate that cognate facilitation depends not only on cross-linguistic overlap but also on the input modality, with orthographic processing showing a greater susceptibility to cross-language influence.

With respect to the sentence-level processing, while some studies demonstrate that cognate facilitation persists even in highly constraining sentences (e.g., Libben & Titone, 2009; Van Assche et al., 2011), others report attenuation or absence of such effects under strong contextual constraints (e.g., Schwartz & Kroll, 2006; Van Hell & De Groot, 2008). Further evidence comes from reading research: in an eye-tracking study of English L1 sentence reading, Titone et al. (2011) found cognate facilitation among highly proficient English–French bilinguals on early eye-movement measures (first fixation duration, gaze duration). These early effects suggested automatic activation of the French cognate forms even in high-constraint English contexts. However, later-stage measures (go-past time, total reading time), which reflect semantic integration, showed reduced cognate facilitation under strong contextual bias. Because the study examined L1 reading of written words, the results cannot be straightforwardly generalized to L2 spoken comprehension. Nonetheless, they indicate that contextual predictability does not eliminate early cross-language activation but may dampen its persistence during later processing stages. Taken together, discrepancies across studies likely arise from differences in semantic constraint, task demands (e.g., eye-tracking vs. lexical decision), the nature of the cognates tested (identical vs. non-identical, high vs. low frequency), presentation modality (written vs. spoken), and psychotypological factors, i.e., bilinguals' subjective perceptions of linguistic distance between their languages (Kellerman, 1979). Speakers of typologically close language pairs may exhibit stronger cognate effects, as perceived similarity increases the likelihood of cross-linguistic activation and transfer (see also Odlin, 1989).

To sum up, cognates provide key insights into the architecture of the bilingual lexicon and the mechanisms of cross-linguistic influence. Their consistent facilitation across tasks supports models of non-selective lexical access, in which lexical candidates from both languages are activated in parallel during word recognition (Dijkstra & Van Heuven, 2002). At the same time, variability in cognate effects across contexts demonstrates that cross-language activation is not fixed but dynamically modulated by factors such as semantic constraint, task demands, L2 proficiency, modality of presentation, and psychotypological distance. Cognates are also particularly relevant for understanding predictive processing in bilinguals: their form–meaning overlap across languages provides a test case for whether anticipatory mechanisms engage lexical representations from both languages or remain restricted to the target language. Although no studies have directly examined cognate pre-activation, early facilitation effects observed during reading (e.g., Libben & Titone, 2009) suggest that cross-linguistic activation occurs rapidly, potentially even before full lexical access. More broadly, research on cognates illustrates how bilingual comprehension is shaped not by a single dominant language, but by the dynamic interplay between two or more linguistic systems.

1.4.2 Language Comprehension in Dyslexia

Developmental dyslexia is a specific learning disorder of neurobiological origin that primarily affects the acquisition of fluent and accurate reading and spelling, which cannot be explained by low intelligence, inadequate instruction, sensory impairment, or lack of motivation. It is a heritable, lifelong condition with an early onset that typically reflects an underlying deficit in the phonological component of language and persists despite adequate educational opportunity (Snowling et al., 2020). A recent comprehensive meta-analysis of 58 studies across multiple languages reported that the global prevalence of developmental dyslexia is approximately 7% among primary school children, with no significant differences in prevalence across writing systems or orthographic depths

(Yang et al., 2022). Orthographic depth refers to the degree of consistency in the mapping between graphemes and phonemes in a given writing system: in transparent orthographies such as Italian or Finnish, the correspondence between letters and sounds is highly consistent and nearly one-to-one, whereas in opaque orthographies such as English or French, these mappings are irregular and variable. Cross-linguistic research confirms that while the surface manifestations of dyslexia differ depending on orthographic transparency, with decoding accuracy being more affected in opaque systems, the underlying cause is universal and rooted in impaired phonological development (Ziegler & Goswami, 2005).

The phonological deficit hypothesis remains the most widely accepted explanatory framework for dyslexia. It posits that individuals with dyslexia possess less precise and less efficiently accessible phonological representations. This phonological deficit is present from the preschool years, long before formal literacy instruction begins, and hinders the acquisition and automatisation of grapheme–phoneme correspondences, thereby impairing fluent reading (Snowling et al., 2020; Vellutino et al., 2004). The deficit manifests most directly in phonological awareness – the ability to detect and manipulate sublexical units such as syllables, onsets, rimes, and phonemes. Phonological awareness is a strong predictor of literacy acquisition across languages and orthographies, persisting into adolescence and adulthood (Carioti et al., 2021; Snowling et al., 2020; Vellutino et al., 2004; Ziegler & Goswami, 2005).

Beyond phonological awareness, individuals with dyslexia often show difficulties in lexical access – the efficiency of retrieving stored phonological representations from memory (Boets et al., 2013; Ramus & Szenkovits, 2008). This difficulty is commonly assessed through rapid automatized naming (RAN) tasks, which measure the efficiency of retrieving phonological labels for familiar visual stimuli such as letters, digits, colors, or objects (Denckla & Rudel, 1976; Norton & Wolf, 2012). Deficits in RAN performance are consistently observed in children and adults with dyslexia and reflect reduced processing efficiency at the interface of visual recognition, phonological retrieval, and motor articulation (Georgiou et al., 2018). Moreover, RAN has been identified as another cognitive marker and predictor of dyslexia across ages and orthographies. While phonological awareness emerges as a stronger predictor specifically in opaque orthographies, RAN remains a robust predictor across all writing systems (Carioti et al., 2021; Reis et al., 2020; Vender & Delfitto, 2024; Vukovic et al., 2004). Another domain that is consistently affected in dyslexia concerns verbal working memory, which supports the temporary storage and manipulation of phonological information during reading and spoken language processing (Just & Carpenter, 1992). It is typically assessed through tasks such as digit span or nonword repetition that tax phonological encoding (Snowling, 2000). Individuals with dyslexia consistently show reduced verbal working memory spans, which negatively affect their ability to maintain phonological codes active during decoding and to integrate information during reading comprehension (Reis et al., 2020).

In addition to phonological awareness and lexical-access deficits, individuals with dyslexia exhibit persistent difficulties in morphological processing. Such difficulties have been documented using both explicit and implicit measures of morphological knowledge across different languages, and are observable from childhood into adulthood (Joanisse et al., 2000; Melloni & Vender, 2022; Schiff & Raveh, 2007; Vender & Delfitto, 2024). Evidence from implicit measures comes from the priming study of Schiff and Raveh (2007), who investigated Hebrew-speaking adults with developmental dyslexia using a word fragment completion task. While typical readers showed robust morphological priming effects, adults with dyslexia did not exhibit facilitation from morphologically related primes. This absence of morphological priming suggests reduced automatic sensitivity to morphemic structure during lexical access, even in tasks that do not require conscious reflection on morphological form.

Evidence from explicit, production-based tasks indicates that morphological difficulties are also observable when individuals with dyslexia are required to actively manipulate morphological structure.

Vender and Delfitto (2024) provided comprehensive evidence for persistent linguistic impairments in Italian adolescents and adults with dyslexia (aged 16–30 years), compared to age-matched typical readers. Their assessment included measures of reading (text, word, and nonword reading), spelling, phonological awareness (nonword repetition, spoonerisms, syllabic and phonemic inversion), morphological awareness (nonword pluralization, past participle inflection, nonword suffix choice), and lexical access (RAN tasks). Results demonstrated pervasive impairments in dyslexia across all measured dimensions. With respect to morphology, adults with dyslexia showed intact performance in some inflectional tasks, such as pluralization of novel nouns, but persistent weaknesses emerged in the inflectional marking of verbs, specifically in forming past participles from pseudowords, as well as in derivational word formation. The study also evaluated the effects of a multi-component web-based personalized intervention combining phonological, orthographic, and morphological training. Following the intervention, adults with dyslexia showed significant improvements in word and text reading, spelling accuracy, and the speed of phonological processing, with smaller gains in nonword reading. These findings underscore that dyslexia is a lifelong condition characterized by persistent deficits not only in literacy skills but also across multiple levels of linguistic competence, including phonological processing, morphological awareness, and lexical retrieval. Importantly, the study also demonstrated that targeted interventions can lead to measurable improvements in adult literacy performance.

Beyond phonological and morphological awareness, individuals with dyslexia experience persistent difficulties in morphosyntactic and grammatical processing more broadly. These difficulties are particularly evident in constructions involving syntactic movement, such as object clitic pronouns, wh-questions, relative clauses, as well as passive constructions. All of these structures place substantial demands on working memory and require maintaining morphosyntactic features across long-distance dependencies. Difficulties have also been reported for sentential negation

(Vender & Delfitto, 2010). Evidence for such morphosyntactic difficulties has been documented in both children (Arosio et al., 2016; Cardinaletti et al., 2022; Guasti et al., 2015; Zachou et al., 2013) and adults (Cardinaletti & Volpato, 2015; Marotta, 2022; Stella & Engelhardt, 2019; Wiseheart et al., 2009).

Evidence that morphosyntactic difficulties persist well into adulthood, even among individuals with high levels of education, comes from Marotta (2022), who used the Italian adaptation of the Test for Reception of Grammar (TROG-2_I) to assess morphosyntactic competence in Italian university students with dyslexia (aged 19–24 years). In this test, participants hear a sentence and must select the corresponding picture among four alternatives, with each block targeting a specific grammatical construction (e.g., passives, relative clauses, coordination, negation, prepositional structures). For instance, in Block T assessing embedded object relative clauses, participants heard sentences like *L'uomo che l'elefante guarda sta mangiando* ('The man that the elephant is looking at is eating') and had to select the image showing the correct thematic role assignment. Compared to age-matched controls, adults with dyslexia showed significantly lower overall performance, with particular difficulties in identifying syntactic subjects in null-subject sentences, reconstructing syntactic sequences with intervening embedded elements between subject and verb, and comprehending relative clauses, particularly object relatives (Marotta, 2022).

Similar conclusions arise from studies of Italian adolescents with dyslexia. Cardinaletti et al. (2022) used a sentence repetition task and an elicited production task to investigate complex structures derived by syntactic movement (clefts, long-distance wh-questions, left-dislocations with clitic resumption, and relative clauses) in high-school students with and without dyslexia. In the sentence repetition task, participants heard complex sentences and were asked to repeat them verbatim. The performance on sentence repetition tasks serves as an indicator of an individual's grammatical ability, as accurate repetition typically suggests that the structure belongs to the participant's grammatical competence (Armon-Lotem & Marinis, 2015).

In the elicited production task, participants viewed images of two referents and were asked to instruct a blindfolded puppet to touch one of them, a context designed to elicit restrictive relative clauses for disambiguation. In repetition, participants with dyslexia produced significantly fewer correct relative clauses than age-matched peers. Specifically, adolescents with dyslexia correctly repeated relative clauses only 22.5–37% of the time, depending on age, whereas typically developing peers showed substantially higher accuracy (higher than 50%). Crucially, experimental sentences containing complex syntax were repeated with lower accuracy than control sentences of identical length, suggesting that the observed difficulties stem from syntactic complexity rather than memory limitations alone. In elicited production, dyslexic participants showed marked difficulties with genitive and prepositional relative clauses, which are characteristic of formal registers and are typically learned through reading. Individuals with dyslexia often replaced the formal genitive relative pronoun *il cui* ('whose') with the colloquial complementizer *che* ('that'), as in **Tocca il papà che il figlio gioca a calcio* ('Touch the dad that the son plays soccer') instead of the target *Tocca il papà il cui figlio gioca a calcio* ('Touch the dad whose son plays soccer'). They also omitted the article component of the prepositional genitive construction *al cui*, producing structures such as **La mamma bacia la bambina a cui il fratello piacciono le tigri* ('The mother kisses the girl to whom the brother "likes" the tigers') instead of the target *La mamma bacia la bambina al cui fratello piacciono le tigri* ('The mother kisses the girl whose brother likes tigers'). Finally, they produced incorrect prepositional forms, for example **Tocca il tetto il cui spazzacamino scende* ('Touch the roof whose chimney sweep descends') instead of the target *Tocca il tetto dal quale scende lo spazzacamino* ('Touch the roof from which the chimney sweep descends'). These findings indicate that processing grammatical features across long-distance dependencies remains vulnerable in dyslexia well into adolescence.

Converging evidence from online measures comes from eye-tracking studies of sentence reading. Stella and Engelhardt (2019) demonstrated that adults with dyslexia show significantly longer reading times at the disambiguating region of garden-path sentences (see more details about the garden-path effect in Section 1.1.2) and lower comprehension accuracy compared to typical readers, with working memory mediating the comprehension deficit. Importantly, the online processing difficulty, indexed by total reading times at the disambiguating verb, remained significant even after variance in working memory was accounted for, pointing to a more specific syntactic integration difficulty in dyslexia beyond what general cognitive limitations can explain.

Although dyslexia is typically diagnosed using reading-related tasks, its phonological deficits can also be observed in spoken language comprehension. Desroches et al. (2006), for instance, used the visual world paradigm to examine auditory word recognition in children with dyslexia. Children heard audio instructions to look at specific objects (e.g., "Now look at the candle") while viewing a display containing four objects (see Figure 1.9 for a display example used by the authors). When presented with displays including a target (e.g., *candle*) alongside cohort competitors (e.g., *candy*) or rhyme competitors (e.g., *sandal*), typically developing children showed interference from both competitor types, whereas dyslexic children were insensitive to rhyme competition. The authors concluded that typical readers categorize auditory stimuli on both their segmental (cohort) and suprasegmental (rhymes) properties, while children with dyslexia are sensitive only to segmental information, despite normal performance on explicit rhyme awareness tests.

Other work has confirmed that spoken comprehension in dyslexia is affected when phonological processing is taxed more heavily. Robertson and Joanisse (2010) examined sentence comprehension in school-age children with dyslexia compared to age-matched controls and children with specific language impairment. The researchers used a sentence–picture matching task where children heard sentences and selected the corresponding picture from four options. The study manipulated two key factors: syntactic complexity (comparing canonical sentences like *"The man is pointing at the boy"* with noncanonical ones like *"The man is pointed at by the boy"*) and working

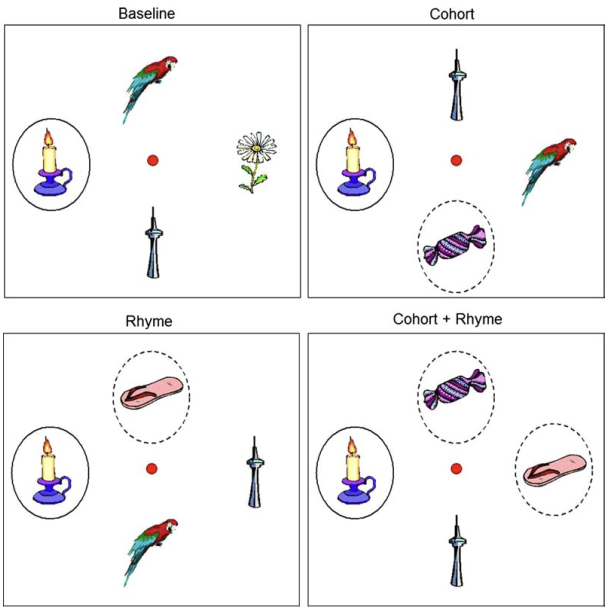


Figure 1.9: Example of a visual display used in Desroches et al. (2006). *Note.* From Desroches, A. S., Joanisse, M. F., & Robertson, E. K. (2006). Specific phonological impairments in dyslexia revealed by eyetracking. *Cognition, 100*(1), B32–B42. © 2005 Elsevier. Reproduced with permission.

memory load by varying when pictures appeared relative to the spoken sentence. Under minimal working memory demands (when pictures were displayed before the sentence onset), children with dyslexia performed as well as controls on both sentence types. Notably, a language-impaired group showed significant comprehension deficits even under these minimal demands, confirming that the relatively preserved performance in dyslexia was not simply due to task insensitivity. However, when memory load increased by presenting pictures only after the sentence ended, all children (including controls) showed greater difficulty with noncanonical sentences, suggesting that syntactic processing becomes challenging for everyone under increased demand. A different, more specific pattern emerged when working memory was taxed even further by adding a 3-second delay between the sentence and the pictures. In this most demanding condition, dyslexic children showed significantly greater difficulty with noncanonical sentences compared to canonical ones, while age-matched controls no longer showed this difference. The authors suggest that age-matched controls may have used their stronger phonological short-term memory to rehearse the sentence during the delay, facilitating syntactic parsing, whereas dyslexic children's weaker storage capacity prevented such processing. This deficit was particularly pronounced for longer sentences, which placed heavier demands on phonological short-term memory.

More recently, Schwarz et al. (2024) demonstrated that phonological deficits also shape how dyslexic adults process familiar words in real-time. Dyslexic participants showed weaker sensitivity both to neighborhood density, the extent to which a word has many similar-sounding neighbors differing by a single phoneme (e.g., *cat*, *cap*, *can*), and to word frequency, suggesting less precise and less efficiently accessed phonological representations. This was evident not only in their behavior but also in the brain: magnetoencephalography revealed reduced activity in the left superior temporal gyrus, a region central to speech perception, between 200 and 500 ms after word onset, the time window when phonological and lexical information is normally integrated. Importantly, these processing differences were observed in the auditory modality during passive listening, confirming that the deficits reflect core language processing difficulties rather than reading-specific problems.

Together, these findings underscore that dyslexia is not confined to difficulties with printed text. Rather, it reflects a deficit in the phonological system that also affects how spoken words are represented, accessed, and integrated during everyday language comprehension. Importantly, these difficulties are not restricted to childhood but persist into adulthood, even among individuals who have achieved compensatory reading skills.

Predictive processing in dyslexia. The studies reviewed above show that dyslexia can sometimes affect online spoken word recognition, with individuals displaying difficulties, especially under increased memory or processing demands. These findings raise the question of whether such difficulties also extend to predictive processing, the ability to anticipate upcoming linguistic input before it is encountered. Prediction is thought to facilitate comprehension by giving listeners a "head start" on word recognition and integration, but if phonological access is impaired in dyslexia, the ability to pre-activate upcoming words and integrate contextual cues during real-time comprehension may be limited.

Importantly, the investigation of predictive processing in dyslexia was directly motivated by earlier work demonstrating that literacy level more broadly modulates anticipatory spoken language processing. Mishra et al. (2012) showed that Indian low literates, unlike high literates, failed to anticipate upcoming target objects in a visual world paradigm even when the sentence context strongly supported prediction, shifting their gaze to the correct referent only well after noun onset. This finding was interpreted as evidence that reading proficiency fine-tunes predictive mechanisms (Huettig & Mishra, 2014). Converging evidence came from Mani and Huettig (2014), who found that word reading scores predicted anticipatory eye movements in eight-year-old children at the cusp of literacy acquisition, independently of spoken language proficiency. Together, these results suggested that it is specifically the acquisition of reading, rather than spoken language experience alone, that sharpens the lexical representations needed for efficient prediction.

Against this background, Huettig and Brouwer (2015) tested whether similar delays would be observed in adults with dyslexia, who typically have substantially reduced reading experience relative to typical readers. They provided one of the first demonstrations that prediction during spoken language processing is impaired in individuals with dyslexia. The authors tested Dutch adults with and without dyslexia in two visual world eye-tracking experiments. The aim of Experiment 1 was to establish whether individuals with dyslexia show the same basic language-mediated eye movement patterns as controls. Both groups were able to use the unfolding speech signal to direct attention to relevant objects. However, crucial group differences emerged in Experiment 2, when grammatical gender on the article served as a predictive cue. Dutch nouns are marked for gender: common-gender nouns take the article *de* and neuter-gender nouns take *het*. Hearing the article, therefore, immediately restricts the set of possible upcoming nouns. For example, on one trial, the target object *piano* was a common-gender noun (*de piano*), while the three unrelated distractors were neuter – *het varken* ('pig'), *het bord* ('plate'), and *het papier* ('paper'). The spoken instruction was *Kijk naar de afgebeelde piano* ('Look at the displayed piano'). The word *afgebeelde* ('displayed') was inserted between article and noun to ensure that participants had enough time to anticipate the target, with the noun onset occurring on average 2009 ms after article onset. Control participants rapidly used the gender cue to anticipate the upcoming referent, shifting their gaze to the correct object well before the noun was spoken. Adults with dyslexia, by contrast, showed a significant delay in anticipating the target. Moreover, the degree of delay correlated with participants' word reading scores, such that poorer readers were also less efficient predictors. Huettig and Brouwer concluded that literacy experience plays a critical role in fine-tuning predictive mechanisms in spoken language. Because adults with dyslexia typically have less reading experience, they may rely less efficiently on grammatical cues such as gender during real-time comprehension, resulting in delayed anticipatory eye movements.

A related study by Engelhardt et al. (2021) investigated prediction in dyslexia using a speeded plausibility judgment version of the cloze task. Adults with and without dyslexia read sentences that varied along two dimensions: (1) sentence constraint (whether the sentence context strongly limited possible continuations or left many plausible options) and (2) cloze probability of the final word. In the high-cloze continuation condition, the sentence ended with the most predictable completion (e.g., "The ruby was so big it looked like a ... cherry"), whereas the low-cloze condition used a less common but still plausible ending (*tomato*). In the anomalous condition, the sentence ended with a semantically incongruous word (*jacket*). Each type of continuation appeared in both high-constraint sentences, where only a few endings were plausible, and low-constraint sentences, where multiple alternatives fit the context (see Figure 1.10 for example stimuli). Participants judged whether each sentence "made sense" as quickly as possible, and reaction times to the sentence-final word were taken as the key measure.

The results showed that both groups were faster for highly predictable completions in high-constraint contexts, demonstrating sensitivity to cloze probability. However, a group difference emerged in low-constraint contexts: while controls still showed faster responses to high-cloze relative to low-cloze endings, individuals with dyslexia did not differentiate between them, suggesting reduced ability in using weaker contextual constraints. Overall, adults with dyslexia were also slower across conditions. The authors interpreted these findings as evidence that adults with dyslexia engage in less incremental semantic processing, i.e., they process the sentence in a shallower way until the final word is encountered, rather than pre-activating likely continuations as the context unfolds. This difficulty was most apparent in low-constraint contexts, where prediction requires maintaining and integrating multiple potential continuations. Engelhardt et al. (2021) noted that the underlying cause remains uncertain. They suggested that limitations in working memory capacity could play a role, given that maintaining and integrating multiple potential continuations places high demands on memory resources, but this factor was not directly assessed in their study. Instead, the authors proposed it as one possible explanation alongside differences in linguistic experience and reading practice. This interpretation aligns with the account offered by Huettig and Brouwer (2015), who argued that literacy experience fine-tunes predictive mechanisms in spoken language comprehension. Although Engelhardt et al. (2021) interpret their findings as evidence of predictive processing, it is worth noting that their measure (reaction times to the sentence-final word) indexes post-lexical facilitation rather than anticipatory activation per se. As the authors acknowledge, they conceptualize facilitated integration and prediction as "two sides of the same coin," meaning that faster responses to predictable endings reflect the effects of prediction.

Working memory capacity is another domain-general factor increasingly linked to predictive processing. Research on neurotypical adults shows that verbal working memory (VWM) enhances both the efficiency and timing of anticipatory language comprehension. Huettig and Janse (2016) found

	High	Low	Anomalous
Low Constraint			
The ruby was so big it looked like a ______.	Cherry (0.14)	Tomato (0.07)	Jacket
I don't know why he didn't take his ______.	Medicine (0.14)	Umbrella (0.09)	Pavement
They went to the rear of the long ________.	Queue (0.15)	Train (0.06)	Nails
Hank reached into his pocket to get the ___.	Money (0.30)	Change (0.09)	Shade
The hunter shot and killed a large _______.	Deer (0.36)	Lion (0.06)	Wind
High Constraint			
The sail got loose so they tightened the __.	Rope (0.54)	Mast (0.06)	Idea
Yesterday, they canoed down the _______.	River (0.81)	Amazon (0.01)	Woods
The ship disappeared into the thick _____.	Fog (0.89)	Mist (0.10)	Cat
At night, the old woman locked the _____.	Doors (0.94)	House (0.03)	Feast
Her job was easy most of the _________.	Time (0.99)	Way (0.01)	Hair

Figure 1.10: Example of sentence stimuli. *Note.* From Engelhardt, P. E., Yuen, M. K. Y., Kenning, E. A., Filipović, L. (2021). Are linguistic prediction deficits characteristic of adults with dyslexia? *Brain Sciences, 11*(1), 59.

that both working memory and processing speed independently predicted anticipatory eye movements during spoken language comprehension in a Dutch grammatical gender prediction task similar to that used in Huettig and Brouwer's (2015) dyslexia study. Using a sample of 105 participants aged 32–77, the authors measured working memory via multiple tasks (nonword repetition, backward digit span, and Corsi block tapping) and processing speed via digit–symbol substitution and letter comparison tasks. Multiple regression analyses revealed that enhanced working memory abilities and faster processing speed each predicted more efficient anticipatory eye movements. Similarly, as discussed earlier (see page 43), Li and Qu (2023) demonstrated that VWM capacity modulated the time course of semantic prediction in Mandarin Chinese, with high-span participants initiating semantic prediction approximately 500 ms earlier than low-span participants. These studies did not examine individuals with dyslexia. Nevertheless, given that VWM is consistently impaired in dyslexia (e.g., Reis et al., 2020; Smith-Spark et al., 2003), reduced working memory capacity may plausibly contribute to delayed prediction in this population, potentially interacting with factors such as limited literacy experience and shallower semantic processing.

Taken together, the available studies suggest that individuals with dyslexia are capable of predictive processing, but that their anticipatory mechanisms are slower and less efficient, particularly in contexts with weaker constraints. In Huettig and Brouwer's (2015) study, this manifested as a delay in using grammatical gender cues to anticipate referents in spoken comprehension, while Engelhardt et al. (2021) found reduced facilitation for less predictable but contextually appropriate continuations during sentence listening. Both studies pointed to limited reading experience as a likely factor constraining prediction in dyslexia, with Engelhardt et al. further emphasizing that this may result in less incremental, shallower semantic processing where individuals wait until the final word is encountered rather than pre-activating alternatives as the sentence unfolds. Moreover, evidence on verbal working memory further suggests that domain-general capacity limitations may modulate the timing of prediction, raising the possibility that such constraints also play a role in the slowed anticipatory processing observed in dyslexia.

With only a few studies directly addressing predictive processing in dyslexia, the evidence base remains too limited to draw definitive conclusions. The present study therefore directly examines whether individuals with dyslexia use grammatical gender information to anticipate upcoming referents, whether they pre-activate semantic and phonological form information of highly predictable words during spoken language comprehension, and whether prediction is modulated by individual differences in reading experience, language proficiency, and cognitive abilities (e.g., processing speed, executive control, and attention) directly comparing to the typical reader population. The study aims to contribute to the existing research and clarify whether predictive differences in dyslexia reflect limitations in the use of specific cues or more general constraints on anticipatory processing.

1.5 The Tyrolean-speaking Community in South Tyrol

1.5.1 Geographic and Demographic Context

The Autonomous Province of Bolzano-South Tyrol (German: Autonome Provinz Bozen - Südtirol; Italian: Provincia Autonoma di Bolzano - Alto Adige) (see Figure 1.11) represents a unique case in Italy as the only multilingual territory where German speakers constitute the demographic majority following the region's annexation to Italy after World War I. According to the 2024 language group census, among Italian citizens resident in South Tyrol, 68.61% belong to the German-speaking language group[6], 26.98% to the Italian-speaking group, and 4.41% to the Ladin-speaking group (ASTAT, 2024).

Figure 1.11: Map of the Autonomous Province of Bolzano–South Tyrol, Italy. *Note.* From Ding, L., Xiao, G., Calvanese, D., et al. (2021). Consistency assessment for open geodata integration: An ontology-based approach. *GeoInformatica, 25*, 733–758. © Springer Nature. Reproduced with permission.

[6] In the South Tyrolean context, "German-speaking language group" refers to the legal-administrative affiliation declared in the census (*deutsche Sprachgruppe* / *gruppo linguistico tedesco*; ASTAT, 2024). This categorization does not necessarily reflect actual language use: members of this group typically speak Tyrolean dialects in daily life, while Standard German is reserved for formal, written, and educational domains (see Section 1.5.2 below for a more detailed discussion).

Moreover, the territorial distribution of these language groups reveals pronounced spatial segregation patterns. Casalicchio (2024) distinguishes three areas of the region based on the main languages spoken: urban areas (mainly Bolzano, Merano, and Bressanone), The South Tyrolean Unterland (Italian: Bassa Atesina, the southernmost part of the region) together with Ladin-speaking valleys, and the rest of the territory. In urban areas, both Italian and Tyrolean coexist, with a majority of the population exhibiting low to intermediate proficiency in their second language (Casalicchio, 2024). Approximately 70% of the region's speakers of Italian reside here. The Bassa Atesina and Ladin valleys are characterized by the use of multiple languages, including Italian, local German dialects, Ladin, and a Romance dialect (Central Trentino), with Standard German present primarily in formal and written domains. The remaining territory in South Tyrol is populated predominantly by members of the German-speaking language group, who typically use Tyrolean dialects (Casalicchio, 2024).

1.5.2 Tyrolean and German in South Tyrol: Linguistic Classification, Diglossia, and Multilingualism

The linguistic situation of the region is best captured by the concept of diglossia. In its classic formulation, diglossia refers to the functional specialization of two language varieties, with one serving as the High (H) variety in formal, institutional, and written domains, and the other as the Low (L) variety in everyday oral interaction (Ferguson, 2000, as cited in Dal Negro, 2010). In South Tyrol, Standard German fulfills the role of the H variety, used in education, administration, and written communication, while Tyrolean dialects serve as the L variety, governing informal spoken communication.

Tyrolean dialects show considerable internal diatopic variation. Lanthaler (2001, cited in Ciccolone 2010) describes this as a "dialectal continuum" (*Dialektkontinuum*) linking local village varieties, unified valley-level dialects (*Taldialekte*), and ultimately a regional *Umgangssprache* (common language) that unites the entire region. From a broader perspective, South Tyrolean dialects are part of the Southern Bavarian group and are continuous with Tyrolean dialects spoken in Austrian Tyrol. They are generally divided into three subgroups: Western Tyrolean (with Alemannic influence, especially in the Vinschgau Valley — *Val Venosta*), Central Tyrolean (including Merano, Bolzano, and Bressanone), and Eastern Tyrolean (covering the Puster Valley — *Val Pusteria* and parts of eastern Austria) (Casalicchio, 2024).

Tyrolean dialect and Standard German differ systematically across all levels of linguistic structure. Phonologically, for example, Southern Bavarian varieties exhibit unique features such as initial laryngeal neutralization patterns absent from Standard German (Vietti et al., 2018). Morphologically, Tyrolean shows extensive case syncretism, particularly in Vinschgau dialects, where nominative, accusative, and dative cases collapse into single forms – a phenomenon documented by the AlpiLinK research project (AlpiLinK, 2023). Lexically, Tyrolean incorporates Italian and Ladin loanwords absent from Standard German, reflecting centuries of contact-induced change (Casalicchio, 2024). These differences explain why Standard German functions for many Tyrolean speakers almost like a second language despite its status as their primary written language acquired through formal education.

The sociolinguistic consequences of this configuration are profound. South Tyrolean speakers often associate Standard German with education, institutions, or tourists from Germany. As a result, some members of the German-speaking group report feelings of linguistic insecurity or inferiority in relation to Standard German speakers from Germany (Leonardi, 2021). As documented by Leonardi (2020, p. 91), South Tyrolean speakers themselves acknowledge that they "hardly ever practice the real Standard German" in spoken contexts, relying instead on dialectal varieties for everyday communication. This creates a distinctive profile of functional multilingualism, where speakers continuously navigate between closely related but systematically distinct varieties of German while simultaneously maintaining competence in Italian.

1.5.3 Sociolinguistic Context and Language Contact

The South Tyrolean sociolinguistic situation has been characterized from multiple perspectives. Some scholars emphasize institutionalized separation between language groups (Vettori et al., 2021), manifested through divided school systems according to language of instruction and employment requirements of bilingual or trilingual competence for public administration positions. Dal Negro (2017) characterizes this configuration as bilinguismo bi-comunitario (bi-community bilingualism), arguing that rather than a single bilingual community, South Tyrol comprises two communities with asymmetric patterns of language contact. Through analysis of sociolinguistic questionnaires and spontaneous speech data from speakers in The South Tyrolean Unterland (Bassa Atesina), Dal Negro demonstrates that Italian is present in virtually all linguistic repertoires in the area, including those of Tyrolean-speakers, while the reverse is not true: many speakers of Italian maintain largely monolingual repertoires. Her empirical analysis reveals implicational hierarchies in which the presence of Tyrolean in a speaker's repertoire always implies the presence of Italian, whereas Italian does not presuppose Tyrolean. This asymmetry also manifests in contact-induced phenomena: analysis of naturally occurring dialogue shows borrowing patterns and code-switching that reveal Italian exerting greater structural influence on Tyrolean varieties than vice versa, establishing a hierarchy of Italian > German > Ladin in terms of penetration into their individual repertoires (Dal Negro, 2017).

As mentioned previously, the German-speaking community in South Tyrol is traditionally diglossic: Tyrolean dialects function as the everyday code, while Standard German is used in formal and written domains. By contrast, the Italian-speaking community, largely shaped by twentieth-century migration, has undergone dialect loss and now relies mainly on a regional variety of Italian (Dal Negro, 2023). This asymmetry extends to patterns of contact and intergroup communication. Italian speakers generally have little access to Tyrolean dialects, since Standard German, which they learn at school as an L2, is not the medium of informal interaction. At the same time, German speakers can more readily use Italian, especially in urban areas where Italian dominates public communication. The widespread use of Tyrolean dialect variants thus presents a significant obstacle to intergroup communication, particularly for Italian speakers (Dal Negro, 2023; Vettori et al., 2012). As Dal Negro (2023, p. 82) observes, "German speakers usually bear the 'burden' of bilingualism much more than Italian native speakers do, and in the case of outgroup communication, the switch to Italian is usually the unmarked choice."

Despite official bilingualism policies requiring second-language instruction from primary through secondary education, the sociolinguistic reality challenges straightforward assumptions about German–Italian bilingual competence among residents (Vettori et al., 2012). The combination of diglossia within the German-speaking community, asymmetric repertoire patterns, and limited access to Tyrolean dialects for speakers of Italian creates a complex multilingual repertoire that shapes the nature and outcomes of language contact in the region.

1.6 Summary

This chapter has brought together theoretical, empirical, and sociolinguistic perspectives to establish the foundations for the empirical work presented in this dissertation. A recurring theme throughout the review is that spoken language comprehension is a fundamentally interactive and predictive process. Contemporary models of lexical access and sentence processing, developed from earlier bottom-up accounts to the newer interactive and constraint-based frameworks, emphasize parallel activation, probabilistic weighting, and the role of contextual information in comprehension.

Within this theoretical landscape, prediction has emerged as one of the central mechanisms of language processing. Across studies, a consistent pattern has been observed: semantic features are readily pre-activated, morphosyntactic cues such as grammatical gender or case can guide expectations under suitable conditions, and phonological form prediction, although demonstrable, tends to be less robust, shorter-lived, and highly constrained by task and context. Such findings indicate that prediction is not an all-or-nothing phenomenon, but a graded process in which the ease and timing of pre-activation differ systematically across linguistic domains.

Research with second-language speakers shows that while anticipatory processing is not restricted to the L1, its reliability depends on proficiency, frequency of exposure, and cross-linguistic influence. Semantic cues are typically used most consistently, while morphosyntactic and phonological cues often generate weaker, delayed, or more variable effects across L2 speakers. Similarly, studies on dyslexia reveal that although anticipatory processing is possible, predictive processing tends to be less efficient. For example, delays in the use of morphosyntactic cues, such as processing grammatical gender (Huettig & Brouwer, 2015) and reduced facilitation from contextual constraint (Engelhardt et al., 2021) show that prediction in dyslexia often unfolds more slowly or less strongly than in typical readers. While the mechanisms underlying this pattern remain unclear, characteristics commonly associated with dyslexia, such as phonological difficulties, slower lexical access, reduced verbal working memory, and limited reading experience, could plausibly shape the dynamics of predictive processing. Finally, I introduced the Tyrolean-speaking population of South Tyrol. The region is characterized by a form of bi-community bilingualism, where German-speaking and Italian-speaking groups coexist in a context of institutional separation, diglossia within the German-speaking community, and highly uneven opportunities for naturalistic cross-linguistic interaction.

Taken together, the literature underscores both the strength of evidence for prediction as a common feature of spoken language comprehension and its variability across linguistic levels, populations, and contexts. These insights directly motivate the empirical questions addressed in the remainder of this dissertation. The following chapters investigate how semantic, morphosyntactic, and phonological information is (pre-)activated in real time, and how these mechanisms unfold in different populations, including Italian and English L1 speakers, L2 speakers, and individuals with dyslexia.

Chapter 2

Experiment 1a: Contextual and Phonological Form (Pre-)Activation in Italian

As discussed in Chapter 1, comprehenders can use semantic, morphosyntactic, and, in some cases, phonological cues to anticipate upcoming linguistic material during spoken sentence comprehension (see Sections 1.2 – 1.3). However, the evidence reviewed there also shows that phonological prediction remains the most contested domain: some studies report pre-activation of phonological form for highly predictable words (DeLong et al., 2005; Ito et al., 2018, 2020; Li & Qu, 2023; Li et al., 2022; Martin et al., 2013), whereas others fail to find reliable effects (Ito & Husband, 2017; Ito & Sakai, 2021; Ito et al., 2017b; Nieuwland et al., 2018) (see Section 1.3.3). The mixed findings suggest that phonological pre-activation may be more fragile, more resource-demanding, or more sensitive to methodological choices than semantic or morphosyntactic prediction. To date, no study has tested phonological prediction in adults with dyslexia, and research on phonological prediction in L2 comprehension remains limited. Moreover, Chapter 1 underscored the need for sensitive, time-resolved measures when investigating prediction. While reaction-time studies (e.g., Engelhardt et al., 2021) have contributed valuable insights, they index facilitation at or after the target word and thus cannot definitively establish whether pre-activation occurred. Eye-movement measures within the visual world paradigm offer a more direct way to capture anticipatory activation with greater temporal precision, as they register rapid gaze shifts to relevant objects even before the target word is heard.

These insights motivate the experimental design adopted in this chapter. By manipulating the informational cues available in the visual scenes and in the linguistic input, the experiments isolate whether listeners rely solely on contextual prediction of the target or whether they additionally pre-activate phonological form. Comparing the anticipatory eye-movement patterns across populations provides a way to determine not only whether prediction occurs, but also how early, how strongly, and at which representational level prediction operates in individuals with different linguistic and cognitive profiles.

Building on this, the present study examines contextual and phonological prediction in Italian among three populations: (i) Italian L1 speakers, (ii) Tyrolean speakers of Italian as an L2, and (iii) Italian L1 adults with developmental dyslexia. The research focuses on the following questions:

1. Are phonological features of an upcoming highly predictable word retrieved before its acoustic onset during spoken sentence comprehension?

2. Do adults with developmental dyslexia differ from typical readers in their anticipatory spoken language processing?

3. Does predictive processing during spoken sentence comprehension differ between L1 and L2 comprehenders?

4. How do patterns of predictive processing compare across the three populations, and what do these comparisons reveal about the mechanisms underlying anticipatory language comprehension?

2.1 Method

2.1.1 Participants

The experiment began with a pilot run involving 12 participants, followed by recruitment of 57 adult participants for the main study. Participants were recruited in South Tyrol through multiple channels, including university mailing lists, advertisements posted on campus bulletin boards, and word-of-mouth referrals within the local community and were reimbursed 8 euros for their participation. All participants gave informed consent, and the study was approved by the University of Verona Ethics Committee – CARP (Comitato di Approvazione della Ricerca sulla Persona).

Participants were divided into three groups: (1) Italian L1 speakers with typical reading development, (2) Italian L1 speakers with a certified dyslexia diagnosis, and (3) Tyrolean L1 speakers without dyslexia who had acquired Italian as a second language (L2). In line with the definition adopted in Section 1.4.1, L2 speakers were understood as individuals who learned the target language after their first-acquired language(s). For the Tyrolean group, eligibility required having grown up with Tyrolean as the sole home language in early childhood, to exclude simultaneous bilinguals. This requirement was made explicit in the recruitment materials (advertising posters and emails) and was verified through participants' self-reports in the Language Experience and Proficiency Questionnaire (LEAP-Q; Marian et al., 2007, see Section 2.1.2.1 for a full description of the instrument), where they indicated the language(s) spoken with parents during childhood. For the dyslexia group, participants were required to provide official documentation of their diagnosis.

Following data cleaning, three participants' data (one from each group) were excluded from the analysis due to insufficient valid trials (less than 30 trials, which fell below 1.5 SD from the mean number of valid trials; see Section 2.2.3 for more information on data cleaning procedures). The final dataset comprised 54 participants across three groups (see Table 2.1).

Table 2.1: *Demographic characteristics of participants by group*

Characteristic	Italian L1		Tyrolean L1
	Typical	Dyslexia	
Sample size (n)	23	12	19
Gender (f/m)	14/9	8/4	12/7
Age (years)	23.1 (4.7)	25.6 (8.5)	26.0 (8.8)
Education level (% secondary/higher)	87/13	75/25	74/26

Note. Values show Mean (Standard Deviation).

Italian L1 typical readers included 23 participants with no reported language disorders, while the dyslexia group consisted of 12 participants with official diagnoses. The Tyrolean group included 19 L1 Tyrolean speakers with Italian as L2. Italian L1 participants reported Italian as their only home language and did not report knowledge of Tyrolean. The groups were comparable in age (Italian L1 typical: M = 23.1 years, SD = 4.7; Italian L1 dyslexia: M = 25.6 years, SD = 8.5; Tyrolean L1: M = 26.0 years, SD = 8.8) and showed balanced gender distributions across groups.

2.1.2 Materials

The study employed the Visual World Paradigm. Participants viewed four black-and-white pictures on the screen while listening to pre-recorded sentences, and their eye movements were recorded. They were told that some sentences would name one of the four objects, whereas others would not, and that their task was simply to look at the picture corresponding to the mentioned object as quickly as possible. In trials where none of the objects were named, they were free to look anywhere on the screen. The design enables the examination of whether listeners pre-activate visual representations of highly predictable words, along with their phonological competitors, based on sentence context.

2.1.2.1 Preliminary Measures

To characterize participants' linguistic background and reading-related skills, a set of assessments was administered. These measures served both to document participants' language experience and to ensure accurate classification of individuals with and without dyslexia.

Language Experience and Proficiency Questionnaire (LEAP-Q). Participants completed the Language Experience and Proficiency Questionnaire (LEAP-Q; Marian et al., 2007), administered online. The LEAP-Q is a commonly used self-report instrument designed to systematically assess multilingual experience. It includes both open-ended and scaled questions, with ratings typically provided on 6-point or percentage scales.

Participants reported broad measures of language dominance by indicating which language they considered their strongest overall, and daily exposure by estimating the percentage of time they used each language in an average day. They also provided language-specific information, including the age of acquisition (when they first started learning each language) and the cumulative length of immersion in environments where the language was spoken (e.g., in the family, at school or work environments, or years spent in a country where the language is spoken). Finally, they self-rated their proficiency in speaking, listening, reading, and writing for each language by responding to the question "How would you rate your ability in [skill]?" on a 1–6 scale, where 1 indicated beginner ability and 6 indicated native-like competence. The results are presented in Section 2.2.

Reading skills assessment. All participants with dyslexia were required to provide documentation of their official diagnosis prior to testing. To ensure accurate group classification and validate dyslexia diagnoses, participants' word and non-word reading skills were assessed using standardized, language-appropriate instruments.

Italian L1 participants (both typical readers and those with dyslexia) completed excerpts from the "Nuova batteria per studenti universitari e adulti LSC-SUA" (Montesano et al., 2020), a standardized assessment designed specifically for university students and adults. This battery was selected because it provides age-appropriate norms for comprehensive evaluation of both lexical (word reading) and sublexical (pseudoword reading) processing routes of the adult population tested in this study. In the word reading task, participants read aloud lists of Italian words of varying frequency and length. They were asked to read the words as fast as they can but trying to avoid mistakes. Reading performance was assessed following the scoring criteria specified in the LSC-SUA manual (Montesano et al., 2020). Performance was measured by reading accuracy (number of errors), reading time (total time to complete the list), and reading rate (syllables per second). See Section 2.2.2 for more details regarding the scoring. The pseudoword reading task followed the same procedure. Participants read aloud pronounceable pseudowords (e.g., *areco, viaca, netro*) that followed Italian phonotactic rules but were not real words. This task assesses phonological decoding abilities independent of lexical knowledge. Accuracy and speed measures were computed using the same operational definitions as in the word reading task. It should be noted that upon inspection, some pseudoword items were found to have limited lexical status (e.g., *areco* as a commercial brand name, *netro* as a toponym). However, these items were retained as presented in the standardized instrument to maintain test validity and comparability with normative data. This highlights the importance of thorough lexical verification when developing pseudoword stimuli, particularly as specialized terms, brand names, and regional vocabulary may not be captured in standard dictionaries. Additionally, periodic verification of existing batteries is recommended to ensure pseudoword items have not entered common usage.

All Italian reading tasks were scored by the author, a non-native Italian speaker with C1 certification, who was trained prior to data collection by a native Italian speaker with graduate-level training in linguistics and prior experience administering and scoring the same battery. Scoring was conducted by the author as a single rater and strictly adhered to the predefined criteria of the LSC-SUA manual across all participants. All scoring was carried out within a restricted time window following data collection, using a fixed set of criteria, in order to minimize intra-rater variability over time.

No standardized reading assessments exist specifically for Tyrolean, which lacks a standardized orthography. For this reason, Tyrolean L1 participants completed the "Salzburger Lese- und Rechtschreibtest II" (Kristina Moll & Karin Landerl, 2014), a standardized German reading assessment. While this test was developed and validated on Standard German speakers, it was considered appropriate given that Tyrolean speakers typically acquire literacy in Standard German alongside Italian through formal schooling. Moreover, the pseudoword reading component, which is the primary indicator of dyslexia, is not sensitive to differences in lexical knowledge, as it involves reading pronounceable but meaningless strings that rely exclusively on sublexical decoding processes.

In the word reading subtest, participants read aloud a list of German words of increasing difficulty, ranging from simple (e.g., Maus [mouse]) to complex words (e.g., Kartoffelschale [potato peel]). Participants were given one minute to read as many words as they could. In the pseudoword reading subtest, participants read German pseudowords (e.g., *oser, gleru, fraulog*) following German phonotactic patterns. The same performance measures were applied as in the word reading subtest. Because the author is not a speaker of German, scoring for this task was delegated to a native German speaker, who annotated errors and omissions in accordance with the guidelines of the test battery. Based on this annotation, performance was quantified as the number of correctly read words in one minute and the percentage of errors, calculated by the researcher. Although different raters were involved for Italian and German tasks, in both cases, scoring procedures were entirely determined by the manuals of the respective standardized tests and applied uniformly across all participants within each language.

2.1.2.2 Sentence Stimuli and Experimental Conditions

The experiment included 42 sentences (see Appendix A.1) that were grammatically correct and semantically plausible. Half of the sentences (experimental sentences) were designed to elicit predictive processing and contained a highly predictable target word (mean cloze probability = 92.4%; see Section 2.1.2.3 for details on the norming procedures) that always came last in the sentence (e.g., *I bambini giocano con la sabbia e costruiscono castelli* [The children play with sand and build castles]). The other half were fillers (e.g., *Tutto il giorno mia nonna gira per casa e cerca i suoi pantaloni* [All day long, my grandmother walks around the house looking for her trousers]) which were always paired with the same image sets used for the experimental sentences (see Figure 2.1). These sentences included semantically non-constraining verbs (e.g., *cercare* [to look for]) that do not strongly restrict the semantic category of their direct object and, unlike experimental sentences, had non-predictive contexts where all four objects on the screen (e.g., *giornali* [newspapers], *occhiali* [glasses], *piatti* [plates], *fiori* [flowers]) were possible continuations of a sentence. Thus, each picture set was seen twice: once in an experimental trial and once in a filler trial. To avoid immediate repetition effects, the experimental and filler sentences that shared the same image set were separated by at least five intervening trials, while maintaining sufficient flexibility for list randomization. Larger minimum distances would have unnecessarily constrained the trial order (e.g., by promoting clustering of filler trials toward later list positions). Four extra sentences representing each experimental condition and filler type were included as practice trials. The presentation order of sentences was fully randomized, as was the spatial arrangement of images on the screen for each trial.

All the sentences followed a similar grammatical structure, containing a first clause that sets up a context, followed by a second clause introducing a related event. The two clauses are connected by conjunctions like *e* [and]. The target was always the last word of the sentence. In the majority of the stimulus sentences (39 out of 41 sentences), the subject remained consistent across the two clauses (e.g., *Laura è molto disordinata e sul divano lascia una forchetta* [Laura is very messy and leaves a fork on the couch]). In two sentences, the subject changed either to a new explicit agent (*La mamma lava i piatti e la bambina disegna un topo* [The mother washes the dishes

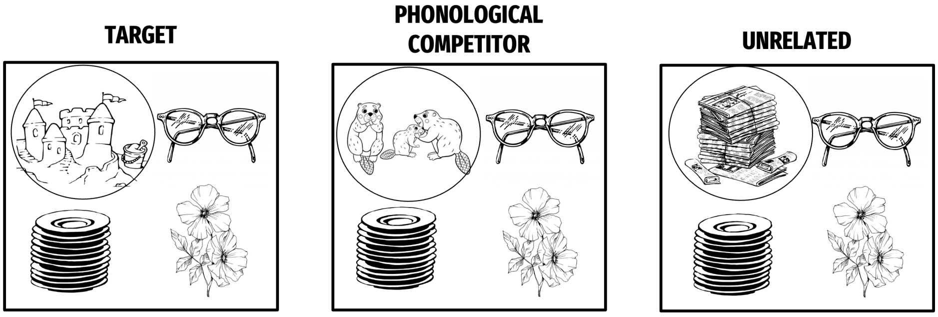


Figure 2.1: Experimental conditions for a sentence: *I bambini giocano con la sabbia e costruiscono castelli* [The children play with the sand and build castles]. In this example, the target object is *castelli* [castles], the phonological competitor is *castori* [beavers], and the unrelated item is *giornali* [newspapers] along with three fixed distractors (*occhiali* [glasses], *fiori*[flowers], *piatti* [plates]). The same visual display was also used for the corresponding filler sentence *Tutto il giorno mia nonna gira per casa e cerca i suoi pantaloni* [All day long, my grandmother walks around the house looking for her trousers].

and the little girl draws a mouse]) or to an implicit plural agent (*Una mattina Guido accende la radio e parlano di un tostapane* [One morning Guido turns on the radio and they're talking about a toaster]). All sentences were written in the present tense. The sentences consisted of a mean of 10.5 words (SD = 1.2, range = 8 – 13 words) or 21.7 syllables (SD = 1.6, range = 19 – 25 syllables), and their mean duration was 4.3 s. The sentences were recorded by a female native speaker of Italian originally from the Veneto region (Northern Italy), who had completed formal training in Italian diction. All recordings were made in a quiet laboratory environment using a Shure PGA48 dynamic microphone connected to an M-Audio Air 192/8 audio interface, with Audacity software. Recordings were digitized at a sampling rate of 44.1 kHz (16-bit WAV). The speaker read sentences at a speed of approximately 5 syllables per second with neutral intonation and without highlighting target words. As in all spoken-language comprehension experiments, participants were expected to become progressively familiar with the speaker's voice over the course of the experiment. However, trial order was fully randomized across participants, ensuring that any effects of voice familiarization were distributed evenly across conditions and could not systematically bias predictability effects. All pictures were black and white line drawings in order to prevent shifts of attention to objects that might be stored with a specific surface color (Huettig & McQueen, 2007). Due to usage restrictions, the visual stimuli are not publicly available, but the full set of stimuli can be shared upon request.

The experimental sentences were presented in one of three conditions, with the corresponding set of images appropriate for that condition (see Figure 2.1). The three experimental conditions were:

- **Target condition**: the critical word was a highly predictable target (e.g., *castelli* [castles]).
- **Phonological condition**: the critical word was a cohort competitor (e.g., *castori* [beavers]) that overlapped, on average, 2.6 phonemes with the target (SD = 0.7; range = 2–4) and matched it in lexical stress. Because pairs with the same number of shared phonemes can still differ in how much of each word matches, I computed the proportion of overlap: the number of shared phonemes divided by the average number of phonemes in the two words (Simmons & Magnuson, 2018). A higher value indicates a greater degree of overlap. For example, longer pairs such as *arcobaleno* – *armadio* [rainbow – wardrobe] have a lower proportion of overlap (0.20) than shorter pairs such as *casa–cane* [house – dog], with a 0.50 overlap, even though both pairs share two initial phonemes. In the current study, the average proportion of overlap was 0.40 (SD = 0.10; range = 0.20–0.60).
- **Unrelated condition (baseline)**: the critical word (e.g., *giornali* [newspapers]) was neither semantically nor phonologically related to the target, and unlike fillers, no words congruent with the sentence context were displayed.

To ensure that both phonological overlap and speech rate were defined with respect to the acoustic signal actually perceived by listeners, stimulus sentences were transcribed using broad phonetic transcription from the recorded audio rather than derived from orthographic or dictionary forms (see Appendix A.2 for full transcriptions). Following Heselwood's (2013) framework, these transcriptions represent broad, auditorily based phonetic categories capturing segmental properties of the speech signal without encoding fine-grained phonetic detail. This approach addresses a gap in the literature on prediction. As Huettig and Guerra (2019) and Fernandez et al. (2025; 2020) have demonstrated, speech rate significantly impacts predictive processing, yet the majority of visual world studies either fail to report speech rate entirely or describe it only qualitatively (e.g., as "normal" or "slow") without quantitative specification. Even fewer studies explicitly measure speech rate and phonological overlap from actual productions rather than dictionary forms. In their comprehensive review, Fernandez et al. (2020) surveyed 45 visual world studies and found that only three reported quantitative speech rates, while most provided no information about the temporal properties of their stimuli at all. This lack of transparency makes cross-study comparisons difficult and may obscure the role of speech rate as a potential confounding variable in prediction effects. Transcribing from actual utterances is particularly important for cross-linguistic comparisons between a stress-timed language like English (used in Experiment 2) and a syllable-timed language like Italian. While Italian phoneme-to-phone mappings are relatively stable and aligned with orthography, stress-timed languages show greater reduction and variability in unstressed syllables (Dauer, 1983; Gardiner & Deterding, 2025; Setter & Sebina, 2017).

String similarity metrics computed over symbolic segmental representations are widely used in computational phonology and dialectometry to quantify phonological distance. More sophisticated approaches incorporate weighted phonological features (Kondrak, 2003), feature distances induced from variation data (Wieling et al., 2012), or acoustic and articulatory distance measures based on spectral and gestural properties of the speech signal (Mielke, 2012). However, the dominant models of spoken word recognition and lexical access in psycholinguistics operate over discrete segment-based representations rather than continuous acoustic or featural representations (Marslen-Wilson, 1987; McClelland & Elman, 1986; Norris, 1994). String-based similarity measures computed over broad phonetic transcriptions are therefore consistent with the representational grain assumed by these models. Importantly, while the resulting measures remain symbolic in nature, they are derived from transcriptions of the actual recorded utterances rather than from dictionary representations, ensuring that phonological overlap is quantified with respect to the specific productions that listeners heard.

Each participant was randomly assigned to one of the experimental blocks. The condition for each sentence was determined by the assigned block. For example, in Block 1, the sentence *I bambini giocano con la sabbia e costruiscono castelli* [The children play with sand and build castles] was assigned to the target condition. Consequently, participants in Block 1 saw an image set containing the target (*castelli*[castles]) along with three fixed distractors (*occhiali* [glasses], *fiori*[flowers], *piatti* [plates]). In contrast, a participant assigned to Block 2, where this sentence was presented in the phonological condition, heard the same sentence paired with the phonological competitor (*castori* [beavers]) and the same three distractors. Similarly, in Block 3, this sentence was presented as part of the unrelated condition. Each participant encountered each experimental sentence only once, with the condition (and corresponding image set) determined by their assigned block.

All the words representing the objects were taken from "*Il nuovo vocabolario di base della lingua italiana*" (Tullio De Mauro, 2016), which comprises the 7,000 most frequent words in Italian. This ensured that the stimuli were drawn from vocabulary items that are highly frequent in contemporary Italian and therefore very likely to be known by both L1 and L2 speakers. This dictionary categorizes words into three groups according to their frequencies: fundamental (most frequent), highly used (moderately frequent), and highly available (less frequent but familiar). For illustration, examples from the present stimuli include *albero* [tree] and *scala* [stairs] (fundamental), *pentola* [pot] and *formaggio* [cheese] (highly used), and *arcobaleno* [rainbow] and *zanzara* [mosquito] (highly available). Within each visual scene, words were balanced across frequency categories, ensuring that picture sets contained items of comparable lexical frequency. When an exact match was not possible (e.g., if three words were *fundamental* but one was *highly used*), a word from the closest frequency group was chosen instead.

Each visual scene contained four objects. Across experimental conditions, only the target condition presented a sentence that mentioned one of the displayed objects. In the phonological and unrelated conditions, the sentence never referred to any visible object. This means that experimental sentences mentioned a visible object 33% of the time (1 out of 3 conditions). Filler sentences mentioned a visible object 67% of the time, balancing the overall distribution so that 50% of all sentences referred to something in the visual scene.

2.1.2.3 Norming Procedures

To validate the stimuli, two norming studies were carried out: a cloze probability (CP) test and a picture naming task. For these norming studies, 30 native Italian speakers who grew up speaking primarily Italian at home and school were recruited, 50% of whom were from South Tyrol. This regional representation was important because the main experiment was conducted in the Autonomous Province of Bolzano - South Tyrol. Including local speakers ensured that the stimuli were appropriate for the linguistic variety encountered by participants in the experimental context. The participants filled out a completely anonymous online survey divided into two parts.

In the first part of the survey, participants completed a cloze probability (CP) test to assess the predictability of target words. They were asked to complete 25 sentences truncated before the target word with the word they felt best fit the sentence (e.g., *leggi la frase e completa con una parola (più l'articolo, se necessario) che ti sembra adeguata: "Marta prepara la frittata e mette un filo d'olio nel/nell'/nella/nello ...*" [read the sentence and complete it with a word (plus the article, if necessary) that seems appropriate to you: "Marta prepares the omelette and adds a drizzle of oil into the..."]). The relative frequency of each provided word was then calculated (the number of times each word was used relative to the total number of responses for that stem), and this frequency defined the CP for individual words. To be included in the experiment, a sentence needed to have a CP of at least 80%, meaning that at least 80% of participants in the norming survey provided the expected target word as the most likely continuation. The threshold for high cloze probability items varies across studies, typically ranging from 0.67 to 1.0 (Lai et al., 2024; Luke & Christianson, 2016). A cutoff of .80 was chosen in the present study to ensure strong contextual constraint while allowing for natural variability in sentence completion responses. This threshold assumes that when most participants (80% or more) produce the same word, that word is genuinely more predictable in that context for the entire population, rather than being the preferred response for only some participants while others expect different words (Staub et al., 2015).

The second part of the survey involved a picture naming task to check the naming agreement for the depicted objects. Participants were shown 376 pictures of objects to be used as picture stimuli and asked to name them using the first word that came to mind. Each picture needed to be labeled by at least 80% of participants with the expected name.

Initially, seven sentences and sixteen images did not meet the indicated criteria. The stimuli were revised, and a second survey with 17 new participants (24% from South Tyrol) was conducted. Following this, all sentences achieved a CP of 80% or higher, and all pictures met the labeling criteria. For instance, the sentence *Chiudo gli occhi e sento il fastidioso ronzio di un/una...* [I close my eyes and hear the annoying buzz of a/an...] initially yielded a CP of 50%, with responses split between *mosca* [fly] and *zanzara* [mosquito], thus failing to meet the predefined predictability criterion. In the revised version, *La notte sento un fastidioso ronzio e vengo punto/a da una zanzara* [At night, I hear an annoying buzz and get bitten by a mosquito], the CP improved to 100%. The mean CP of the target word in the final set of sentence stimuli was 92.4% (SD = 7.4, range = 80 − 100%). Tyrolean L1 speakers were not included in the norming procedures, as I would be norming Italian stimuli with L2 speakers of Italian rather than establishing baseline predictability values in the target language. The exceptionally high cloze probabilities achieved (mean 92.4%) indicate very strong semantic-contextual constraints that should remain predictable across speaker populations, as high constraint contexts have been shown to benefit both native and L2 speakers (Ma et al., 2015).

2.1.3 Procedure

Upon arrival, participants provided informed consent and completed the background questionnaire and reading assessments described in Section 2.1.2.1. These preliminary tasks were administered prior to the eye-tracking experiment to document participants' linguistic experience and reading-related skills. The eye-tracking session took place in a sound-attenuated booth. The experiment was programmed and presented using OpenSesame (Mathôt et al., 2012). Auditory stimuli were presented as uncompressed waveform audio files (.wav), which were loaded directly into OpenSesame and played back in their original format. Participants' eye movements were recorded using a Gazepoint GP3 HD Eye Tracker sampling at 60 Hz (Gazepoint, 2024). The sampling rate is how many times the eye position is measured per second. In the context of measuring eye fixations during a spoken language comprehension task, a sampling rate of 60 Hz is considered adequate despite being lower than rates typically used for reading research, such as 500 Hz or 1,000 Hz. As Raney et al. (2014) explain, while higher sampling rates yield better temporal accuracy, particularly for capturing the precise timing of saccades, a 60 Hz sampling rate — equivalent to recording eye position every 16.7 milliseconds — still provides sufficient accuracy for studying fixation durations. For spoken language comprehension, where the focus is primarily on fixations rather than rapid saccades, this level of temporal resolution is appropriate (Raney et al., 2014).

Participants listened to the auditory stimuli over Sennheiser HD 599 open-back headphones (frequency response: 12–38,500 Hz; impedance: 50 Ω; sensitivity: 106 dB SPL at 1 kHz). At the beginning of the session, participants were instructed to adjust the output level to a clearly audible setting. Volume adjustments were permitted only before the initial calibration phase and again after the practice trials if needed; the volume then remained fixed throughout the experimental blocks.

Benvenuto/a!

Mentre sentirai una frase, vedrai 4 immagini sullo schermo. A volte (non sempre!) una delle immagini verrà menzionata nella frase. Guardala il più velocemente possibile!

Cerca di mantenere sempre gli occhi sullo schermo. Prima che compaiano le immagini, ci sarà un piccolo punto al centro dello schermo. Ti preghiamo di mantenere gli occhi fissi su questo punto quando appare.

Figure 2.2: Written instructions (Italian original) shown to participants at the beginning of the experiment. English translation: "Welcome! While you listen to a sentence, you will see four pictures on the screen. Sometimes (not always!) one of the pictures will be mentioned in the sentence. Look at it as quickly as possible. Try to keep your eyes on the screen at all times. Before the pictures appear, a small dot will be shown in the centre of the screen. Please keep your eyes fixed on that point when it appears."

Written instructions in Italian appeared on the screen (Figure 2.2) and were also read aloud by the experimenter to ensure comprehension. Participants were instructed to listen carefully to each sentence and keep their eyes on the screen throughout the trial. They were informed that sometimes (but not always) one of the pictures would be mentioned in the sentence, and their task was to look at the mentioned picture as quickly as possible. Participants were given the opportunity to ask questions before the experiment began.

Before data collection, the eye tracker was calibrated to ensure accurate mapping between the participant's gaze position and the corresponding screen coordinates. During calibration, participants fixated a series of points on the screen, allowing the system to compute the difference between eye-movement signals as measured by the eye-tracker and true eye position on the visual space (e.g., Godfroid, 2020). This was done using a nine-point calibration grid (see example in Figure 2.3). Four practice trials preceded the first experimental part, after which participants could ask questions; these were mainly limited to clarifying the instructions and did not reveal

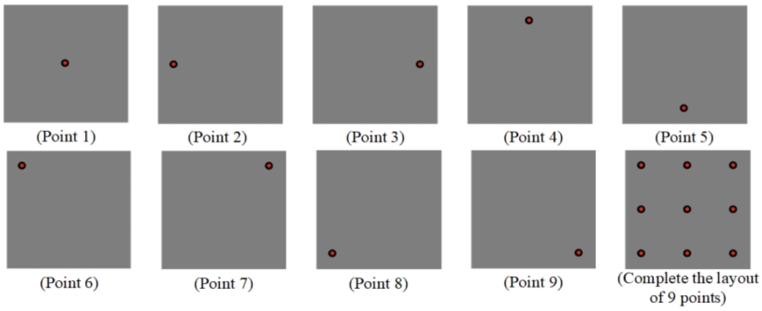


Figure 2.3: Example of a nine-point calibration grid used in the experiment. The fixation target appeared sequentially at the nine positions shown here. *Note.* Reproduced from Lu, S., Li, R., Jiao, J., Kang, J., Zhao, N., & Li, M. (2020). An Eye Gaze Tracking Method of Virtual Reality Headset Using a Single Camera and Multi-light Source. *Journal of Physics: Conference Series, 1518*, 012020, under the terms of the Creative Commons CC-BY 3.0 licence.

additional difficulties. Drift correction was performed at the start of each trial to ensure that the participant's current fixation corresponded accurately to the centre of the screen. During this procedure, the eye tracker continuously sampled gaze position and checked whether the participant maintained fixation within a 50-pixel radius around the screen centre for at least 500 ms. If this criterion was not met within 10 s, the trial was interrupted and a full recalibration procedure was initiated before continuing. After successful drift correction, the audio recording played, and four pictures appeared on the screen at a resolution of 1280 × 1024 pixels. The visual presentation of the picture stimuli began 2000 ms before the acoustic onset of the critical word. An example of the experimental procedure is shown in Figure 2.4. After the end of Experiment 1a, participants took a short break and continued with Experiment 1b (described in Chapter 3).

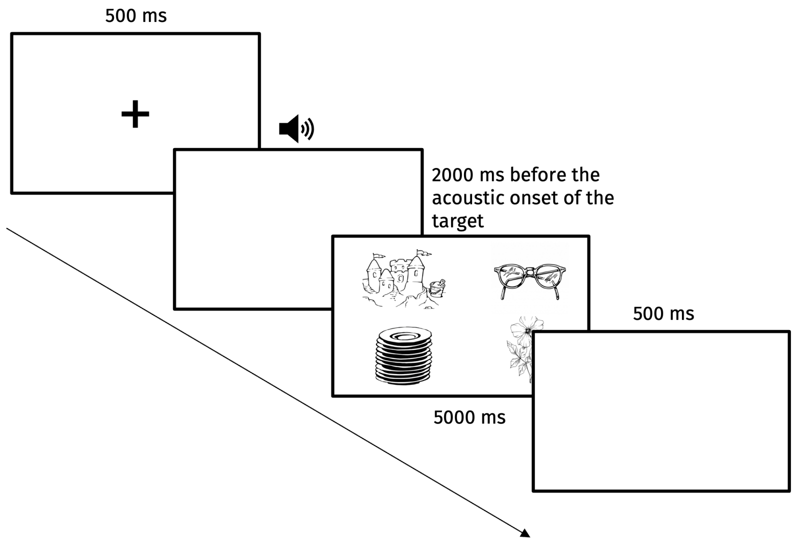


Figure 2.4: An example of the experimental procedure.

2.1.4 Predictions

Based on the theoretical framework and empirical findings reviewed in Chapter 1.3, I formulated the following predictions for the present study.

Contextual prediction across populations. Drawing on extensive evidence that comprehenders routinely anticipate highly predictable targets in strongly constraining contexts (e.g., Altmann & Kamide, 1999; Borovsky et al., 2012; Ito et al., 2018), I expected all three groups to show anticipatory looks to target objects before word onset. Given these findings, I predicted that L2 speakers would show anticipatory target fixations similar to (or with only minor differences from) those of L1 speakers at the group level, though individual differences in language experience (as measured by the LEAP-Q) might correlate with the timing and strength of these effects. For adults with dyslexia, previous research demonstrated reduced efficiency in using contextual cues, specifically in low-constraint contexts (Engelhardt et al., 2021), and delayed anticipatory eye movements in grammatical prediction tasks (Huettig & Pickering, 2019). Given that the present materials featured highly constraining contexts (mean cloze probability = 92.4%), which are more similar to the high-constraint sentences where Engelhardt et al. found preserved (though numerically slower) prediction in dyslexia, I expected anticipatory target fixations to occur in adults with dyslexia, but with a delay in their onset relative to typical readers. This prediction is further motivated by evidence that adults with dyslexia show persistent difficulties in lexical access, as indexed by slower rapid automatized naming performance (Carioti et al., 2021; Vender & Delfitto, 2024), which may delay the time course of predictive eye movements even when contextual constraints are strong enough to support prediction.

Phonological form prediction. If comprehenders pre-activate the phonological form of the expected noun and this activation spreads to phonologically related lexical representations, items that phonologically overlap with the predictable target should attract anticipatory fixations prior to target onset. Whether such effects emerge is an empirical question, given the contradictory findings in the literature. Evidence for phonological form pre-activation during spoken comprehension is considerably more limited and variable than for semantic or morphosyntactic prediction (see Section 1.3.3). Meta-analytic evidence indicates that phonological competitor effects, when detected, are small (approximately 5% increase in looks) and short-lived (roughly 600 ms) (e.g., Ito, 2024; Ito et al., 2018). Accordingly, I predicted at most weak and transient phonological competitor effects in Italian L1 typical readers. For L2 speakers, evidence for phonological prediction is limited (see Section 1.4.1). Studies using the English *a/an* determiner alternation have found that proficiency modulates phonological anticipation, with higher-proficiency L2 speakers showing earlier and stronger effects (Connell et al., 2021). However, research examining whether L2 speakers pre-activate the phonological form of predictable targets strongly enough for competitors to be activated has yielded largely null results. Ito et al. (2018) found no anticipatory looks to phonological competitors in Japanese L1 English L2 speakers, observing competitor fixations only after target onset, consistent with bottom-up priming rather than prediction (see page 61). For adults with dyslexia, I predicted no phonological competitor effects. This prediction was motivated by consistent evidence of phonological processing difficulties in dyslexia, including reduced sensitivity to phonological information during spoken word recognition (Desroches et al., 2006; Schwarz et al., 2024) and impaired phonological awareness across the lifespan (Snowling et al., 2020; Vender & Delfitto, 2024). Although the precise nature of phonological deficits in dyslexia remains debated, whether they reflect underspecified phonological representations or difficulties in accessing otherwise intact representations (see Ramus & Szenkovits, 2008), both accounts predict reduced phonological pre-activation during real-time comprehension.

Comparing populations. Both L2 speakers and adults with dyslexia were expected to show differences in predictive processing compared to L1 typical readers, though potentially for different underlying reasons. Research reviewed in Chapter 1.3 suggests that L2 prediction difficulties may stem from weaker lexical-semantic mappings (Dijkgraaf et al., 2019), reduced automaticity in linguistic processing (Ito & Pickering, 2021), and increased cognitive demands of L2 comprehension (Corps et al., 2023). In contrast, prediction difficulties in dyslexia likely reflect core phonological processing deficits (Snowling et al., 2020), reduced lexical access efficiency indexed by slower rapid automatized naming performance (Vender & Delfitto, 2024), and more limited reading experience (Huettig & Pickering, 2019). These group differences are expected to be quantitative rather than qualitative: all groups should ultimately identify the target, but the onset and magnitude of anticipatory effects may differ (Kaan, 2014; Schlenter, 2023).

2.2 Data Analysis and Results

2.2.1 Language Experience

Table 2.2 presents linguistic background information for all participant groups based on the LEAP-Q. Tyrolean speakers reported first exposure to Italian at a mean age of 5.8 years (SD = 1.8, range = 3 – 10 years), typically coinciding with entry into primary school. All participants indicated that they acquired Tyrolean from birth, confirming sequential rather than

Table 2.2: *Language background characteristics from LEAP-Q assessment*

LEAP-Q Variable	Italian L1		Tyrolean L1
	Typical	Dyslexia	
L1 Language Experience:			
Daily exposure to L1 (%)	67.4 (17.5)	66.0 (19.6)	66.8 (19.9)
L2 (Italian) Language Experience:			
Age of acquisition (years)	–	–	5.8 (1.8)
Daily exposure (%)	–	–	20.1 (10.4)
Italian Academic/Professional Exposure:			
None	–	–	33%
1–5 years	–	–	17%
More than 5 years	–	–	50%
Self-Rated L1 Proficiency:			
Speaking	5.9 (0.4)	5.7 (0.7)	5.7 (0.7)
Listening	5.9 (0.4)	5.7 (0.6)	5.7 (0.7)
Reading	5.8 (0.5)	5.6 (0.9)	5.5 (1.0)
Writing	5.8 (0.6)	5.4 (1.0)	5.3 (1.0)
Average	5.9 (0.4)	5.6 (0.7)	5.6 (0.8)
Self-Rated L2 (Italian) Proficiency:			
Speaking	–	–	4.2 (1.1)
Listening	–	–	4.6 (0.8)
Reading	–	–	4.4 (0.9)
Writing	–	–	3.8 (1.2)
Average	–	–	4.3 (0.9)

Note. Values show Mean (Standard Deviation) unless otherwise specified. L1 = first language; L2 = second language. Italian Academic/Professional Exposure indicates the duration of time spent in schools and/or workplaces where Italian was the primary language of communication. Self-rated proficiency measured on a scale from 1 (lowest) to 6 (highest).

simultaneous bilingual acquisition, with Italian introduced primarily through formal instruction. Tyrolean accounted for an average of 66.8% of daily language use (SD = 19.9%), Italian for 20.1% (SD = 10.4%), and the remaining 13% reflected use of additional languages. Although daily L1 exposure percentages may appear relatively low at first glance, this pattern likely reflects the composition of the sample: the majority of participants (all but two) were students at the Free University of Bozen-Bolzano, where German, Italian, and English are all official languages of instruction. Because participants were required to distribute their overall daily language use such that percentages sum to 100%, exposure to Italian and English in academic contexts naturally reduces the proportion attributed to Tyrolean in the questionnaire. Consistent with the diglossic repertoire of Tyrolean speakers in South Tyrol (see Section 1.5.2), all participants in this group also reported knowledge of Standard German (mean age of acquisition = 5.3 years, SD = 1.4), and of English as an additional language (M = 10.0 years, SD = 2.2). One participant additionally reported Ladin and two reported French, all learned later than Tyrolean and not spoken in the home. Although Tyrolean has no standardized orthography, participants reported their self-rated literacy skills based on Standard German used in education and written communication. All participants had lived in Italy for at least five years at the time of testing. Academic and professional Italian exposure varied across participants: 33% reported no experience in Italian-medium educational or work environments, 17% reported 1-5 years of such experience, and 50% more than 5 years. Self-rated proficiency in Italian averaged 4.3 out of 6 (SD = 0.9) across modalities, indicating an intermediate to high-intermediate level of L2 competence. Participants rated receptive L2 skills (listening M = 4.6; reading M = 4.4) slightly higher than productive skills (speaking M = 4.2; writing M = 3.8). This receptive–productive asymmetry is consistent with patterns documented in L2 research (Dornic & Ekehammar, 1988), particularly in the domain of vocabulary knowledge, where receptive knowledge typically exceeds productive knowledge (Laufer & Paribakht, 1998). Importantly for the present study, which focuses on spoken language comprehension, listening was the highest-rated modality.

All Italian L1 participants, except three who had already completed their studies, were students at the Free University of Bozen-Bolzano. Both typical readers and those with dyslexia reported that Italian was their dominant language. Italian accounted for an average of 67.4% of daily language use in the typical-reader group and 66% in the dyslexia group, with the remainder reflecting other known languages. All additional languages were learned after L1 Italian, and none of the participants were simultaneous bilinguals. All Italian L1 participants reported knowledge of English, and several also indicated French, Standard German, Spanish, or Russian. None reported Tyrolean, which is expected in the context of South Tyrol: as shown by Dal Negro (2017), Italian does not presuppose access to Tyrolean in speakers' repertoires, whereas the reverse implication holds for Tyrolean speakers (see Section 1.5.3). Overall self-rated Italian proficiency on the 1-6 LEAP-Q scale was very high for typical readers (M = 5.9, SD = 0.4) and only slightly lower for participants with dyslexia (M = 5.6, SD = 0.7). Within the dyslexia group, self-rated proficiency was lowest for written modalities (writing: M = 5.4; reading: M = 5.6) compared to oral modalities (speaking and listening: M = 5.7). Compared to typical readers, participants with dyslexia showed lower self-ratings across all modalities, with the largest difference observed for writing (0.4 points) and comparable differences for speaking, listening, and reading (0.2 points each). Notably, even oral modalities (speaking and listening) were rated slightly lower by participants with dyslexia, which may reflect either genuine differences in spoken language processing or the broader impact of dyslexia on linguistic self-perception (see Glazzard, 2010; Humphrey & Mullins, 2002; McNulty, 2003, for discussion of self-concept in dyslexia). These self-ratings likely capture both genuine literacy-related challenges associated with dyslexia and participants' heightened metacognitive awareness of these challenges.

2.2.2 Reading Assessment Results

Reading assessment results confirmed successful group classification into the three participant groups (Italian L1 typical readers, Italian L1 readers with dyslexia, and Tyrolean L1 typical readers). All subcomponents were z-standardized within each language group to ensure comparability across test formats and to allow group-level comparisons of reading performance. A composite reading score was computed for each participant by averaging z-standardized scores from two tasks: word and pseudoword reading. For each language, word and pseudoword reading scores were standardized separately. In the Italian test, reading performance was assessed via a combined measure of speed and accuracy: total reading time, reading rate (syllables per second), and number of errors. Accuracy was calculated by assigning one point per correctly read word or pseudoword without omissions or alterations of its form. Self-corrections and strong hesitations were not counted as errors, as they are considered to affect reading speed rather than accuracy. Examples of errors included segmental deviations that altered the target phonological form (including substitutions, insertions, and deletions). Segmental deviations included, for instance, reading /pɑntoˈmiːma/ as /pɑntonoˈmiːma/ or /maˈnɔpola/ as /manoˈpɔlja/ for words, and reading /tostolaska/ as /toskolaska/ or /ɔmbaraːda/ as /ombrarada/ for pseudowords. Lexical or morphological modifications in real-word reading included productions such as /inʤɛˈɲɛre/ as /indʒeˈrire/, /perˈikolo/ produced as /perikoˈlozo/ and /aˈnarkiːka/ produced as /anarˈkiːa/. Errors in stress placement were counted as errors for real words (e.g., reading /monogaˈmia/ as /monoˈgamja/), with the exception of the words *omega* and *balia*, for which both stress patterns were considered correct as indicated in the test manual. In the German assessment, performance was likewise based on accuracy and speed, but speed was indexed by the number of correctly read items within a fixed 1-minute time window, following the standardized procedure (Kristina Moll & Karin Landerl, 2014). Errors were scored according to the test manual, with one error assigned for each incorrectly read item that was not self-corrected, with a maximum of one error per item regardless of multiple letter errors within that word. Examples of German errors included segmental deviations affecting the target phonological form, such as reading /ˈbratsən/ instead of /ˈbraːtən/ or /ˈʃtit/ instead of /ˈʃnit/ for words, and /svaːku/ instead of /ʃvaːku/ or /maːma/ instead of /moːma/ for pseudowords. In addition, real-word replacements were observed, in which a different existing lexical item was produced, for example reading /tɔpf/ as /kɔpf/ or /ˈʃʊpn/ as /ˈʃnʊpfn/. Finally, morphological alterations in real-word reading included changes in inflectional morphology, such as reading /ˈkɔpfʃmɛɹts/ as /ˈkɔpfʃmɛɹtsən/ or /ˈmɛnʃ/ as /ˈmɛnʃən/.

Descriptive statistics for reading scores by group can be seen in Table 2.3.

Although the two batteries differ in the way timing is implemented, they both assess the integration of speed and accuracy in word and pseudoword reading. Z-standardization within each language group addresses this metric difference by converting raw scores to a common scale that preserves individual differences relative to language-appropriate norms, consistent with cross-linguistic research (e.g., Landerl et al., 2013; Ziegler et al., 2010).

Table 2.3: *Descriptive statistics for reading scores by group*

Group	Mean	SD	SE
L1 (L1 Italian typical readers)	-0.03	0.63	0.13
L2 (L1 Tyrolean typical readers)	0.16	0.46	0.11
L1 Dyslexia (L1 Italian dyslexic readers)	-2.19	1.04	0.30

Note. SD = standard deviation; SE = standard error. Composite reading scores are averaged z-scores of word and pseudoword reading measures.

It should be noted that while specific errors were annotated during scoring, only summary scores (total errors per task) were retained for analysis. This decision reflects the scope of the present study: the reading assessments were not intended to provide a detailed error analysis or to diagnose dyslexia, but rather to serve as a control measure to confirm group classification.

Table 2.4: *Welch ANOVA and Games–Howell post-hoc test results for reading scores across groups*

	Statistic	df		p
Welch ANOVA	$F = 26.9$	$(2, 24.8)$		$< .001$
Games–Howell Post-Hoc Comparisons:				
	Mean Difference	t	df	p
L1 vs. L2	-0.199	-1.17	39.5	.475
L1 vs. Dyslexia	2.16	6.60	15.4	$< .001$
L2 vs. Dyslexia	2.36	7.42	13.8	$< .001$

Note. Games–Howell post-hoc comparisons following Welch ANOVA. Degrees of freedom are adjusted using the Welch–Satterthwaite approximation. p-values are two-tailed.

Composite reading scores showed clear differentiation between groups (see Table 2.3 for descriptive statistics and Figure 2.5 for a visual

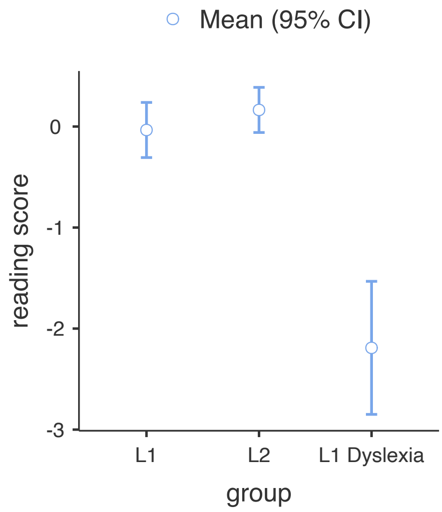


Figure 2.5: Group means with 95% confidence intervals for standardized reading scores across the three participant groups: L1 (Italian typical readers), L2 (L1 Tyrolean, L2 Italian typical readers), and L1 Dyslexia (L1 Italian participants with dyslexia). Scores are z-standardized within each language group; values near 0 indicate average performance, positive values above-average, and negative values below-average. Error bars represent standard errors.

comparison). A Welch ANOVA[1] confirmed a significant main effect of group on standardized reading scores, F(2, 24.8) = 26.9, p < .001 (see Table 2.4). Games–Howell post-hoc comparisons revealed that participants with dyslexia had significantly lower reading scores than both Italian typical readers (mean difference = 2.16, p < .001) and Tyrolean typical readers (mean difference = 2.36, p < .001). The difference between the two typical reader groups was not significant (mean difference = -0.199, p = .475). These results confirm a clear separation of the dyslexia group from both groups of typical readers, whereas no reliable difference emerged between Italian L1 and Tyrolean L1 typical readers.

[1]Welch ANOVA was chosen because the data showed unequal variances across groups, and it is robust to violations of the homogeneity of variance assumption.

2.2.3 Data Cleaning

Gaze data were preprocessed using the eyetracking-R vignette (Dink & Ferguson, 2015). Fixation proportions were coded as looks at critical object, distractors, and "other" (any looks outside the predefined interest areas) for each 50 ms time bin. Trackloss occured when the eye-tracker was unable to reliably record the participant's gaze due to blinks, looks off the screen, or poor gaze location validity. To ensure data quality, trials with a trackloss proportion greater than 27% were removed (210 trials, 8% of total trials). Furthermore, trials in which participants never fixated the target object throughout the trial were excluded (2 trials: 0.08% of all trials). This exclusion was made because a lack of attention to the target, even after the target word had been spoken, suggests the participant may not have been attending to the auditory input or the visual display. While fixation of the target object does not in itself guarantee successful perception or identification of the spoken word, random or accidental fixations are not expected to yield systematic, time-locked gaze patterns aligned with the speech signal. Inference about spoken word processing relies on condition-dependent fixation dynamics across trials and participants, rather than on single-trial fixation outcomes. Additionally, participants who contributed fewer than 1.5 standard deviations below the mean number of valid trials due to excessive trackloss were excluded (three participants, one from each group). High trackloss in these cases was attributed to factors such as excessive head movement, difficulty maintaining fixation on the screen, excessive blinking, or calibration issues that could not be resolved during the session (in one case, related to contact lens reflections).

Accuracy and precision are two central indices of eye-tracking data quality (Godfroid, 2020). Accuracy refers to how close the recorded gaze position is to the participant's true gaze position, whereas precision indicates how stable the gaze signal is when the eyes are fixated on one location. These measures are independent: data can be accurate but imprecise (i.e., correct on average but noisy), or precise but inaccurate (i.e., consistently offset). High-quality eye-tracking data ideally combine both high accuracy and high precision, ensuring that fixation events are correctly identified and reliably assigned to visually distinct regions of interest. In the present study, calibration accuracy was recorded for every participant at the beginning of the session. Across participants, accuracy typically ranged between approximately 0.25° and 0.40° of visual angle (min = 0.17°, max = 0.50°). Precision, quantified as RMS noise, corresponded to roughly 0.3°–0.5° of visual angle, indicating low sample-to-sample variability. In practical terms, this means that the eye tracker placed participants' gaze estimates very close to their actual point of fixation (within $< 1°$ of visual angle), demonstrating both high accuracy and stable measurement. Moreover, these accuracy and precision values are also consistent with validation studies of the Gazepoint GP3 system, which report spatial accuracy between 0.5° and 1° of visual angle (with no reported precision values) (Brand et al., 2021). For comparison, the EyeLink 1000 has a stated accuracy of $< 0.5°$ and precision (RMS) of approximately 0.01°–0.02° (SR Research Ltd., 2005).

2.2.4 Data Analysis Approach

First, to investigate language processing across different population groups, I visualized the data by plotting time-course graphs illustrating the mean fixation proportions of looks to the critical word in each condition over time across all participant groups (Figure 2.6).

To statistically investigate the impact of experimental condition and participant group (L1 vs. L2; typical vs. dyslexic) on the proportion of looks to the critical object over time, I implemented two generalized additive mixed models (GAMMs), each targeting a distinct time window[2]:

- **Anticipatory window:** -1800 ms to 200 ms relative to target word onset, corresponding to the period from picture onset until shortly after the acoustic onset of the target word.
- **Resolution window:** 200 ms to 850 ms relative to target word onset,

[2]Time windows follow standard practices in visual world paradigm studies examining predictive processing (Ito and Knoeferle, 2023; Godfroid, 2020, p. 198).

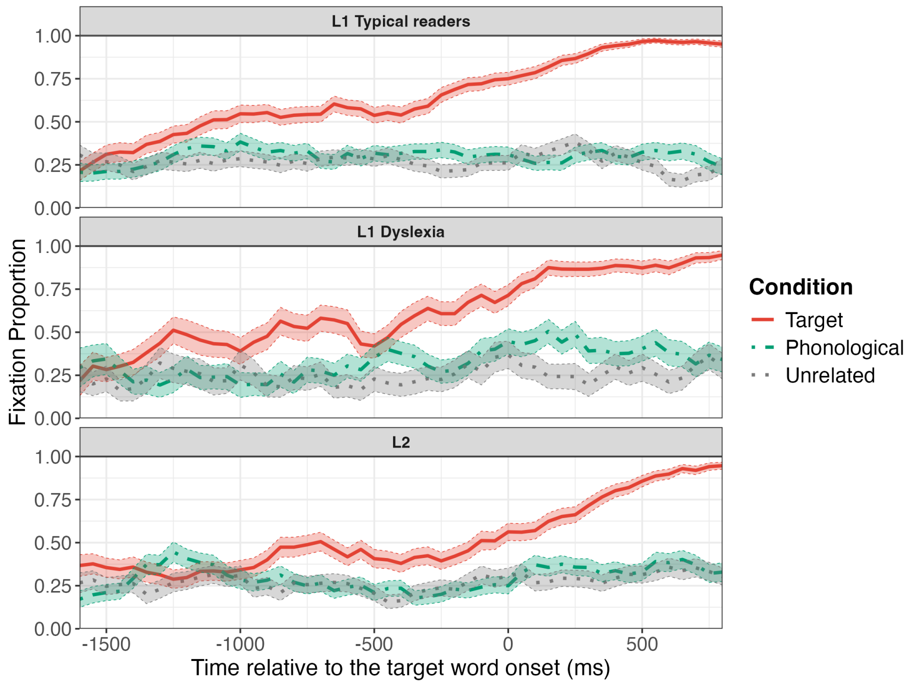


Figure 2.6: Mean fixation proportions on target, phonological competitor, and unrelated (baseline) objects in the L1 (top graph), Dyslexia (middle graph), and L2 (bottom graph) groups. Time 0 ms represents the acoustic onset of the target word. Error bands indicate standard errors.

corresponding to the period during (mean target duration ∼630 ms) and shortly after the auditory presentation of the target word capturing target recognition and integration effects.

Following standard practice in the eye-tracking research, I added 200 ms to these time windows to account for the time required to initiate a saccade (Saslow, 1967). Although 0 ms corresponds to the acoustic onset of the target word, eye movements launched in response to this word typically appear around 200 ms later due to processing and motor delays. As a result, looks directed at the target or competitors within the anticipatory window (–1800 ms to 200 ms) are interpreted as anticipatory, because they occurred before the spoken word could have influenced eye movements.

For both time windows, I fitted GAMMs using the `bam()` function from the `mgcv` package (Wood, 2017). The dependent variable was the empirical logit–transformed proportion of looks to the critical object (`elog`). Gaze data were prepared using the `make_time_sequence_data()` function from the `eyetrackingR` package (Dink & Ferguson, 2015), which structures the data as one observation per participant × item × 50 ms time bin. Within each time bin, the function computes a fixation proportion (`prop`) of gaze samples directed to the area of interest relative to all on-screen samples in that bin (sampled at 60 Hz), rather than by aggregating across participants or items. No prior averaging across participants or items was therefore performed before modelling, and the data retained their full trial-level structure. The model specification was as follows:

```
main <- bam(elog ~
    s(time) + cg +
    s(time, by = cg) +
    s(time, participant, by = cg, bs = "fs") +
    s(time, item,  bs = "fs"),
  data = data, rho = rho1, AR.start = Is_start,
  method = "fREML")
```

Each model included smooth terms for time (measured in milliseconds) and parametric terms for the interaction between experimental condition (target, phonological competitor, baseline) and participant group (L1, L2, Dyslexia), encoded in the factor variable `cg`, which represented the full crossing of condition and group. The `cg` variable was constructed as an ordered factor with treatment (dummy) coding, with `control.L1` (the unrelated condition in the Italian L1 typical-reader group) as the reference level. As a result, the parametric coefficients reported in the model output (see Appendix D) reflect contrasts against this baseline: the Intercept estimates mean fixation in the reference cell, and the remaining coefficients represent differences in mean fixation for each condition–group combination relative to `control.L1`. The difference smooths `s(time, by = cg)` correspondingly capture how the time course of fixations in each condition–group cell departs from the reference trajectory across the time window. Because the data retained their full trial-level structure, random variation across both participants and items was modelled directly through the by-participant and by-item factor smooth random effects included in the GAMM (i.e., `s(time, participant, by = cg, bs = "fs")` and `s(time, item, bs = "fs")`). Item-level random smooths by condition (`s(time, item, by = cg)`) were initially considered but led to overparameterized models without improving model fit or interpretability and were therefore excluded from the reported models. To account for temporal autocorrelation in the data, an AR(1) parameter (ρ) was estimated and incorporated into the models.

To find the best-fit model, my selection process involved two key comparisons. First, to test the contribution of the condition $\times$ group interaction (*cg*), each full model was compared against a corresponding null model that excluded the interaction. Second, models with and without the autocorrelation parameter were compared to evaluate whether including ρ improved model fit. In both cases, comparisons were conducted using the `compareML()` function from the `itsadug` package (Van Rij et al., 2015).

In time series or time-based data, like eye-tracking data with dense time bins, residuals are not independent across time points. Therefore, it is important to address autocorrelation. People typically move their eyes only every 200 – 300 ms despite more frequent eye-tracker sampling (Ito & Knoeferle, 2023). Thus, fixation at time 't' is likely correlated with fixation at time 't–1'. A major advantage of GAMMs is their ability to account for such autocorrelation, thereby reducing the risk of inflated false-positive rates. However, binomial distribution cannot be used in GAMMs to account for autocorrelation, so I used empirical logit (*Elog*) transformed fixation proportions as the dependent variable. The transformation was calculated with a correction factor ($\epsilon = 0.5$) added to both numerator and denominator of the logit equation to handle proportions of zero or one (Dink & Ferguson, 2015).

To probe significant interactions, pairwise smooth comparisons were conducted using functions from the `itsadug` package (e.g., `plot_diff()`, `get_smooths_difference()`) with cluster-based permutation tests. These procedures identify time windows where trajectories significantly diverge, providing the GAMM-appropriate analogue of post-hoc contrasts in linear models (such as `emmeans`). The same approach was applied consistently to all between- and within-group comparisons. For clarity of presentation, results are divided into separate paragraphs by group comparison, but they all stem from this post-hoc procedure. While GAMMs are sensitive to when groups begin to diverge in their fixation trajectories, they do not provide an explicit statistical comparison of divergence onsets between groups. They capture differences in the overall proportion of looks and their time courses, but cannot test whether, for example, one group starts to diverge earlier than another. Therefore, I complemented GAMM analyses with divergence point analysis (DPA), which is explicitly designed to estimate the timing of divergence between conditions and groups, providing confidence intervals for divergence onset and statistical tests of timing differences (described in more detail in the next Section 2.2.5).

2.2.5 Results. Anticipatory Window

A comparison between the constructed GAMM, including the condition $\times$ group interaction, and its corresponding null model (without this interaction) using the `compareML()` function from the `itsadug` package revealed that the interaction significantly improved model fit (ΔfREML = –17,446.110, Δdf = 22.0, $p < .0001$, AIC difference = 35,233.55). This indicates that experimental condition and group jointly influenced anticipatory looking behavior. Incorporating the autocorrelation parameter ($\rho = 0.82$) also substantially improved model fit compared to the model without this parameter (ΔfREML = –15,984.73, Δdf = 0, $p < .0001$, AIC difference = 31,790.18), confirming significant temporal dependencies in the eye-tracking data. Full model output is reported in Appendix D.1. Smooth terms revealed significant differences in anticipatory looking behavior across conditions and groups. As illustrated in Figure 2.7A, the empirical logit trajectories clearly diverge between target and unrelated conditions across all participant groups, providing evidence of strong preactivation of the critical word in the target

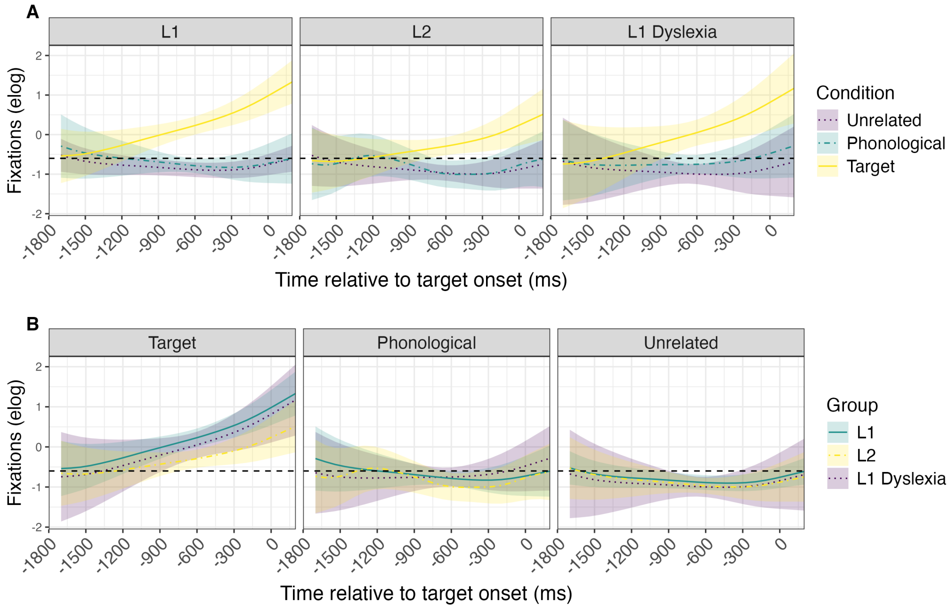


Figure 2.7: GAMM model estimates of looks to the critical object over time. (A) Comparison by condition for each group. (B) Comparison by group for each condition. The x-axis represents time relative to the target word onset (0 ms), and the y-axis shows the proportion of looks directed to the critical word transformed into empirical logit. If the confidence interval (CI) of the graphed smooth doesn't overlap with the dashed chance line, one can infer that the effect is significant and the smooth is performing above random chance. When the CI of two displayed smooths do not intersect at a particular point in time, it suggests that the two smooths are statistically distinct at that moment.

condition.

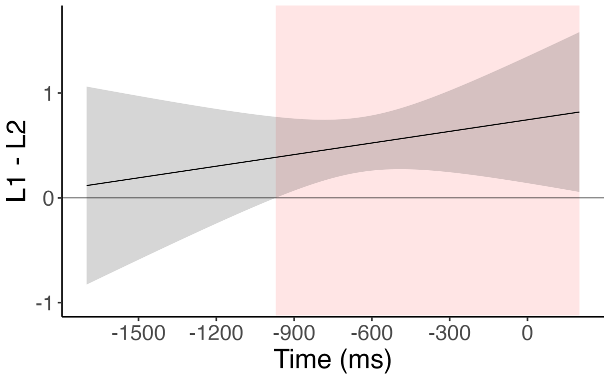
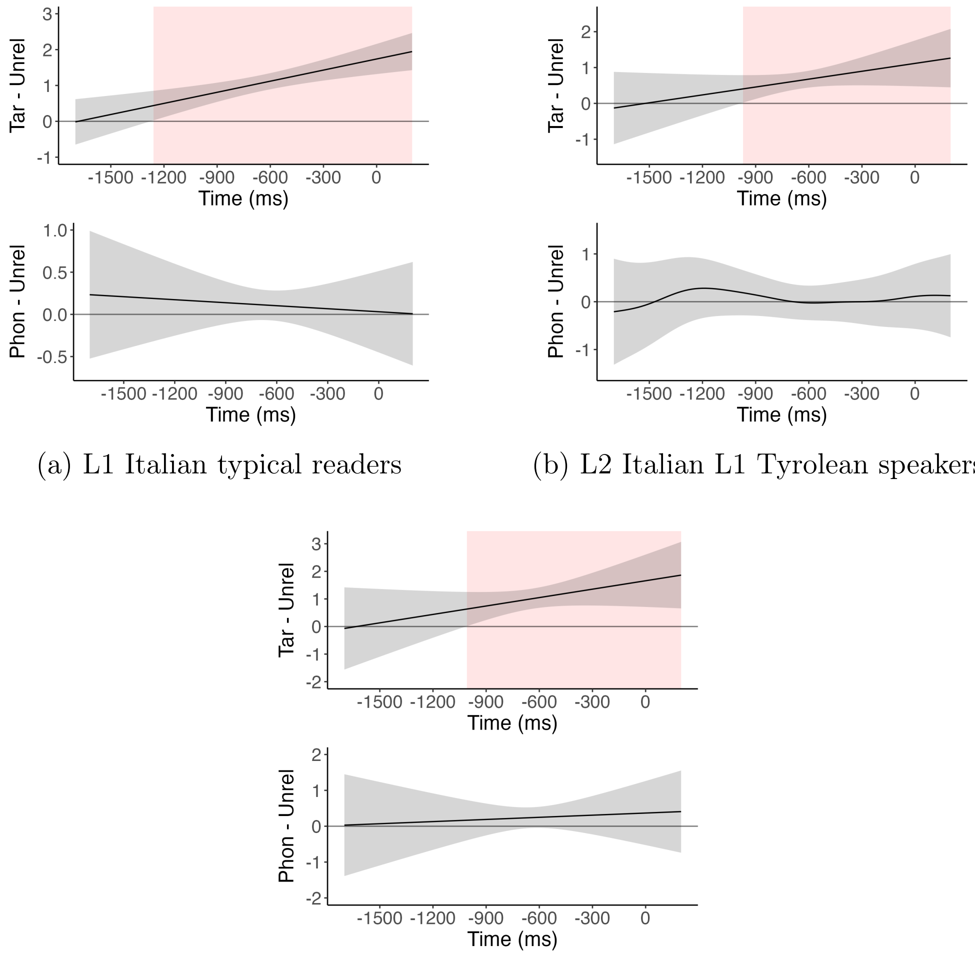


Figure 2.8: Estimated differences in empirical logit-transformed fixation proportions over time in the *target* condition between L1 Italian typical readers and L1 Tyrolean speakers (Italian L2). The x-axis represents time relative to target word onset (0 ms). The solid line represents the estimated mean difference, with positive values indicating higher fixation proportions for L1 Italian typical readers compared to L1 Tyrolean speakers. Shaded bands represent 95% confidence intervals. Pink shading marks time intervals during which group differences are statistically significant.

Group differences in anticipatory processing: L1 vs L2. Despite exhibiting significant anticipatory fixations to the target overall, Tyrolean speakers showed reduced anticipatory engagement compared to Italian speakers, as reflected in a consistently lower proportion of looks to the target in this time window. The difference plot demonstrates that the group effect increased gradually over time, with the confidence interval excluding zero entirely at approximately -971 ms (effect size = 0.603), indicating a significant difference between the groups (Figure 2.8). To estimate the time at which anticipatory effects occurred for L1 and L2 groups separately, I conducted difference smooths analyses comparing target and control conditions within each group. These analyses revealed that, in the L1 group, looks to the critical word in the target condition began to diverge from looks to the critical word in the unrelated condition approximately 1259 ms (effect size = 1.192) before target onset (Figure 2.9a), while in the L2 group, significant divergence occurred later, around 971 ms (effect size = 0.834) before target onset (Figure 2.9b).



(a) L1 Italian typical readers

(b) L2 Italian L1 Tyrolean speakers

(c) Participants with dyslexia

Figure 2.9: Panels show differences in empirical logit-transformed fixation proportions over time between the target and unrelated conditions (upper graph in each panel) and the phonological and unrelated conditions (lower graph in each panel) for (a) L1 Italian typical readers, (b) L2 Italian L1 Tyrolean speakers, and (c) participants with dyslexia.

To statistically evaluate whether the divergence point occurred significantly earlier for Italian speakers compared to Tyrolean speakers, I conducted a divergence point analysis (DPA) using a bootstrapping approach (see Figure 2.10). For this analysis, I ran 2000 bootstrap iterations resampling the data with stratification to maintain the original data structure. For each time point in each resampled dataset, I compared log-transformed fixation proportions between target and control conditions using t-tests. A significant divergence was defined at the first timepoint that initiated a list of four consecutive t-values exceeding 1.96 (equivalent to $p < 0.05$). Thus, for each bootstrap sample, I calculated the divergence point (DP) for each group separately and the difference between these DPs, which allowed me to estimate when each group began predictive looking toward targets with 95% confidence intervals. To determine whether the observed difference in DPs between the two groups was statistically significant, I computed a null distribution by randomly reassigning group labels across 2000 iterations. Then, for each bootstrap sample with reassigned labels, I calculated the temporal difference between the two groups' DPs. The proportion of bootstrap samples where this difference was less than or equal to zero represented the probability of observing a difference as large as or larger than the one obtained under the null hypothesis that there is no difference between groups.

The results showed that Italian L1 speakers began to show preferential looks to the target at a mean onset of -1156 ms (95% CI [-1350, -900] ms), while Tyrolean speakers showed this divergence later, at -855 ms (95% CI [-950, -450] ms). These were similar to the GAMM model predictions (-1259 ms and -971 ms, respectively), indicating a good model fit to the data. The mean difference in divergence points between the groups was approximately 251 ms (95% CI = [0; 750] ms), which reached the conventional threshold for statistical significance ($p = 0.05$). This indicates a trend toward earlier anticipatory processing in L1 compared to L2 speakers. However, given that the confidence interval included zero and that there was overlap in the group-level confidence intervals (L1: [-1350, -900] ms; L2: [-950, -500]

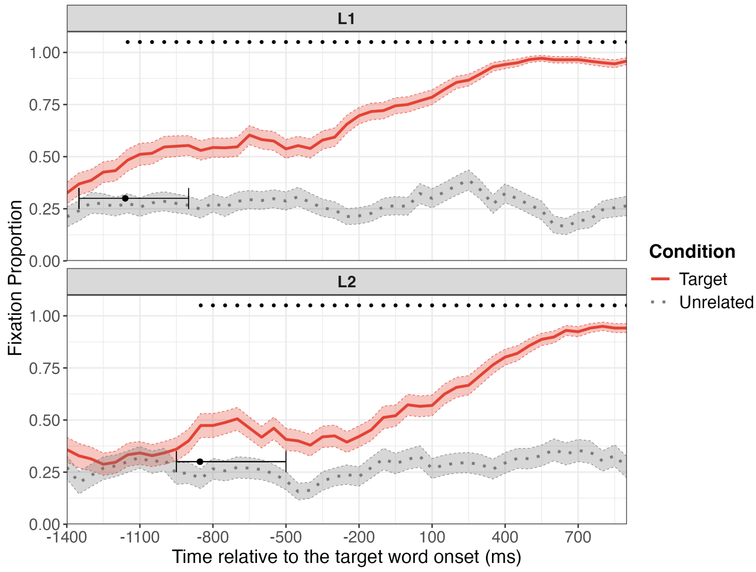


Figure 2.10: Divergence point analysis of mean fixation proportions to the critical word in target vs. unrelated conditions for L1 and L2 groups. Fixation proportions to the critical word are plotted over time relative to the target word onset (0 ms), separately for L1 and L2 groups. The y-axis shows the mean proportion of fixations to the critical word. Error bars represent standard errors. The black dot and accompanying horizontal error bar mark the estimated divergence point and its 95% confidence interval for each group. Black dots above the plots indicate timepoints at which GAMM estimates reveal a statistically significant difference between the target and unrelated conditions.

ms), this effect should be interpreted with caution.

In line with the predictions outlined in Section 2.1.4, which anticipated that individual differences in L2 experience might modulate the timing of anticipatory processing in L2 speakers, I further examined whether variability in divergence point estimates within the Tyrolean group was associated with measures of language experience and proficiency. To this end, Pearson correlations were computed between each participant's mean divergence point in the highly predictable target condition and a set of LEAP-Q variables indexing Italian exposure, self-rated proficiency, age of acquisition, and continued L1 use (see Appendix G for the full correlation table).

Given the limited sample size (n = 17–19 across variables), the present correlational analyses should be interpreted with caution. The smaller sample size for daily Italian exposure (n = 17) reflects missing self-report data for this variable in two participants. With samples of this size, correlation estimates can be unstable. Accordingly, the correlations are interpreted in terms of effect size and direction rather than statistical significance. Three correlations were of low-to-medium magnitude: daily Italian exposure (r = 0.28), self-rated L2 proficiency (r = 0.19), and age of Italian acquisition (r = –0.22). The remaining correlations were negligible (|r| < 0.08).

The direction of the correlations is not straightforwardly interpretable. Since more negative divergence point values reflect earlier anticipatory fixations to the highly predictable target, the positive associations with daily Italian exposure (r = 0.28) and self-rated L2 proficiency (r = 0.19) would suggest that higher exposure and perceived proficiency are weakly associated with later anticipation of contextually predictable words. This pattern seems to run counter to expectations. The negative association with age of Italian acquisition (r = –0.22) would similarly suggest that later learners showed marginally earlier anticipatory fixations, which is equally difficult to interpret. Given the small sample size, however, all three estimates are inherently unstable, which motivates the exploratory analyses described below.

No univariate outliers were identified in divergence point estimates within the Tyrolean group based on standard interquartile range (1.5 × IQR) and z-score (|z| > 2.5) criteria. To further assess whether the observed correlations were driven by individual observations, two complementary analyses were conducted in R. First, diagnostics based on Cook's distance (Belsley et al., 1980) were computed from simple linear regression models predicting divergence point from each language experience variable. Cook's distance is a measure of how much a single participant influences the overall result. For each participant in turn, it asks: "If I removed this person from the dataset, how much would the correlation change?" A participant with a high Cook's distance is therefore an influential observation — not necessarily an outlier in the traditional sense, but someone whose data point has a disproportionate pull on the estimated relationship. A conventional threshold of $4/n$ (where n is the sample size) was used to flag influential observations ($\approx$ 0.24 for analyses with n = 17 and $\approx$ 0.21 for analyses with n = 19). Second, leave-one-out (LOO) correlations were computed by recalculating each correlation repeatedly, removing one participant at a time and estimating the relationship on the remaining sample (Hastie et al., 2009). This gives a direct picture of how stable the correlation estimate is across the sample: a narrow LOO range suggests the correlation is robust, while a wide range suggests it is heavily dependent on specific individuals. For daily Italian exposure (r = 0.28), two influential observations were identified (Cook's D = 1.09 and 0.25). When both influential observations were removed, the correlation increased to r = 0.49. LOO correlations ranged from r = 0.18 to r = 0.57, indicating that the full-sample estimate is not stable and varies substantially depending on which participants are included, although the direction of the association remains consistently positive. For self-rated L2 proficiency (r = 0.19), two mildly influential observations were identified (Cook's D = 0.26 and 0.22). Their removal produced only a modest change in the correlation (r = 0.19 to r = 0.23), and LOO correlations ranged from r = 0.09 to r = 0.32, suggesting that the estimate is relatively stable in direction but remains weak. In contrast, for age of Italian acquisition (r = –0.22), one influential observation was identified (Cook's D = 0.5). Removing this participant reduced the correlation from r = –0.22 to r = –0.03, eliminating the effect. LOO correlations ranged from r = –0.30 to r = –0.03, indicating that the negative correlation is entirely driven by a single observation and cannot be considered a reliable finding.

Taken together, these analyses indicate that the observed correlations are not driven by simple outliers in divergence point estimates, but rather reflect varying degrees of sensitivity to individual observations. While the associations with daily Italian exposure and self-rated L2 proficiency show weak but directionally consistent trends, their magnitude is unstable. The correlation with age of acquisition is not robust and does not provide reliable evidence of a relationship. Importantly, no coherent pattern emerges across predictors. Given the limited sample size, all estimates carry substantial uncertainty, and these findings should be treated as exploratory observations.

Group differences in anticipatory processing: Italian L1 typical readers vs. dyslexia. Within-group analyses revealed that both participants with dyslexia and Italian L1 typical readers showed target preactivation effects. In the dyslexia group, looks to the critical item in the target condition began to diverge from the unrelated condition around 1009 ms before target onset (effect size = 1.248), while in the typical reader group, divergence occurred at 1259 ms before target onset (effect size = 1.192) (Figures 2.9c and 2.9a, respectively).

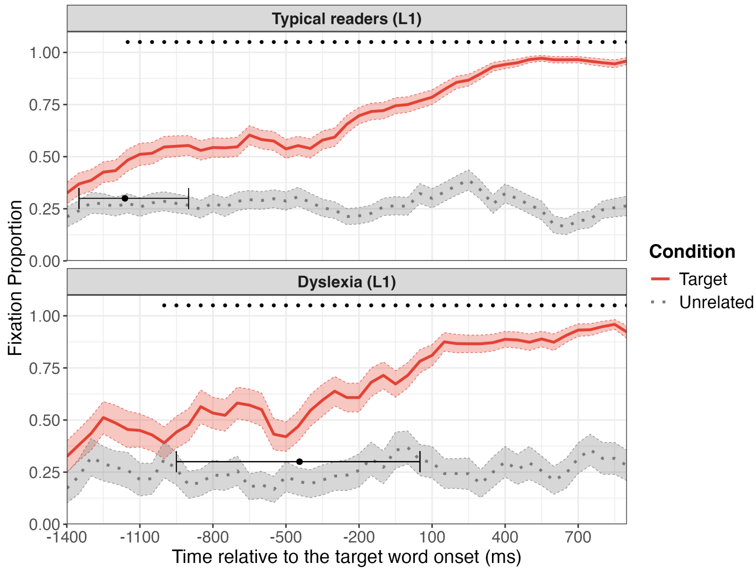


Figure 2.11: Divergence point analysis of mean fixation proportions over time for the target and unrelated conditions. Fixation proportions to the critical word are plotted over time relative to the target word onset (0 ms), separately for typical readers (Control L1) and participants with dyslexia (Dyslexia L1).

However, divergence point analysis detected greater variability in timing across participants (Figure 2.11). According to this analysis, typical readers initiated predictive fixations approximately 1156 ms before target onset (95% CI [-1350, -900]), while participants with dyslexia showed a mean divergence at 441 ms before target onset (95% CI [-950, 50])[3]. Despite a 664 ms

[3]It is important to note that the DPA estimate reflects the average divergence onset across participants, while its confidence interval indicates the plausible range of this mean. By contrast, GAMMs identify the earliest time point at which fixation trajectories differ reliably. As a result, the GAMM estimate typically aligns more closely with the earlier

difference in mean divergence timing, the between-group difference did not reach significance ($p = 0.2$, 95% CI [150, 1300]). It is also worth noting that the confidence intervals in the dyslexia group were substantially wider (-950 to 50 ms) compared to the typical reader group (-1350 to -900 ms), indicating greater between-subject variability.

To further characterize the variability within the dyslexia group, I inspected the distribution of individual mean divergence points (Figure 2.12). These individual estimates represent each participant's mean divergence timing computed directly from their fixation data, and are distinct from the bootstrapped group-level DPA estimate and its confidence interval, which reflects sampling uncertainty around the group mean across 2000 bootstrap resamples[4]. The distribution spanned from –1384 ms to –314 ms (mean = –998 ms, SD = 293 ms, median = –1045 ms, MAD = 271 ms). While visually the distribution appears somewhat spread, with one participant showing a notably later divergence point (–314 ms) relative to the rest of the group, no extreme values were detected based on z-score screening ($|z| > 2.5$), and the proximity of the mean and median suggests the distribution was not substantially driven by extreme observations. Moreover, the null between-group result is unlikely to be an artefact of this participant's influence: removing them would shift the dyslexia group mean toward earlier divergence, thereby reducing the difference between groups. The wide confidence interval in the DPA, therefore, appears to reflect between-subject variability rather than the influence of a small number of atypical cases, while also being expected given the relatively small sample size of the dyslexia group, which limits the precision of the estimate. Importantly, this degree of heterogeneity is consistent with the well-documented individual variability within the dyslexia population, where differences tied to severity are an

end of the DPA confidence interval rather than with its mean. This is why the GAMM estimate of –1009 ms is nearer to the lower bound of the DPA CI (–950 ms) than to the DPA mean (–441 ms). The two methods therefore provide complementary rather than conflicting timing estimates.

[4] In other words, the DPA confidence interval answers the question: *how precisely can we estimate when this group, as a whole, diverges?* The histogram of individual divergence points answers a different question: *how much do individual participants vary in their divergence timing?*

expected feature of the population (e.g., Snowling, 2001; Wagner et al., 2020). Given that no statistical grounds for exclusion were identified, all 12 participants were retained.

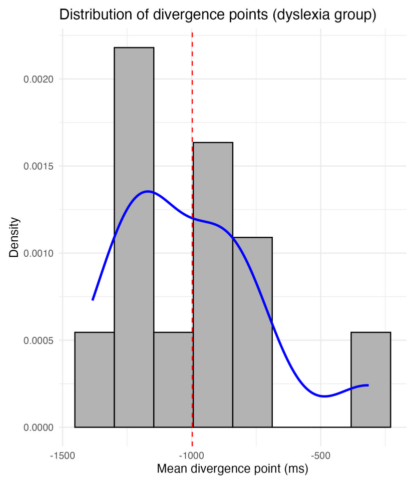


Figure 2.12: Distribution of individual mean divergence points (ms before target onset) in the dyslexia group (n = 12). The red dashed line indicates the group mean (–998 ms). The overlaid density curve (blue) is shown for reference. Individual divergence points represent each participant's mean estimated divergence timing and span from −1384 ms to −314 ms.

The GAMMs difference curves comparing the patterns of anticipatory eye movements directed at the critical object in the target condition between the two groups (Figure 2.13) also showed no significant differences, with confidence intervals including zero throughout the entire time window. This confirms the results of the DPA analysis that both groups demonstrated comparable anticipatory abilities. However, the high individual variability in the dyslexia group (evidenced by the wide confidence intervals) suggests this null result should be interpreted cautiously, as the heterogeneity may mask subgroup differences or reduce power to detect more subtle between-group effects.

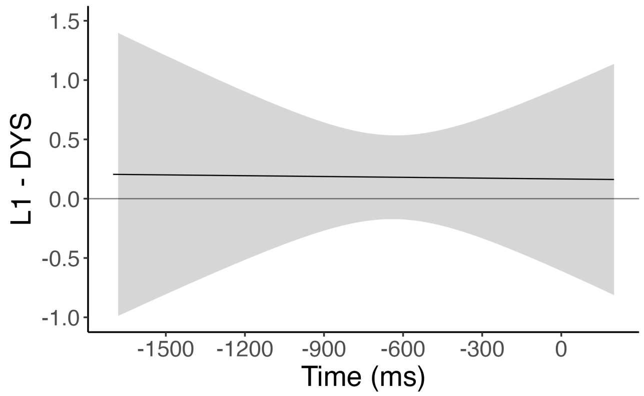


Figure 2.13: Estimated differences in empirical logit-transformed fixation proportions over time in the *target* condition between L1 Italian typical readers and L1 Italian participants with dyslexia. The x-axis represents time relative to target word onset (0 ms).

Group differences in anticipatory processing: L2 vs. dyslexia. Comparisons were conducted also between non-typical groups to see whether L2 typical readers and L1 speakers with dyslexia differ from each other in their anticipatory processing patterns. Figure 2.14 shows the difference in fixation proportions between L2 and dyslexic participants in the target condition across the anticipatory time window. As seen in the figure, the confidence interval includes zero throughout the entire time window, indicating no statistically reliable difference between the groups at any timepoint.

These results are further supported by the divergence point analysis, which also did not detect any significant difference between the groups ($p = 0.4$), with the mean divergence of 464 ms between the groups (CI: -1000, 50) (see Figure 2.15). The substantial overlap in confidence intervals (L2: [-950, -500] ms; Dyslexia: [-950, 50]) confirms that despite their different sources of processing difficulty (second language vs. reading disorder), both groups demonstrate remarkably similar dynamics in their anticipatory language processing.

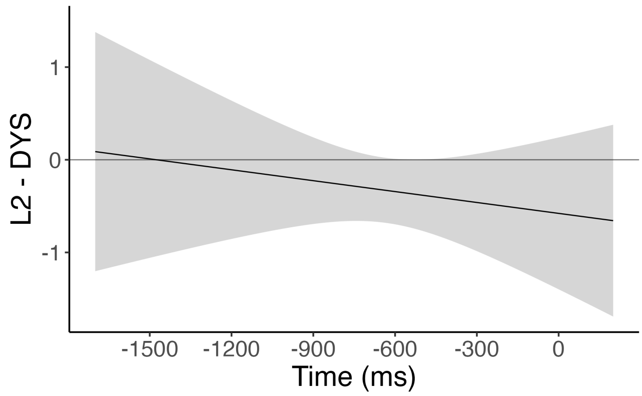


Figure 2.14: Estimated differences in empirical logit-transformed fixation proportions over time in the *target* condition between L1 Tyrolean speakers (Italian L2) and L1 Italian participants with dyslexia. The x-axis represents time relative to target word onset (0 ms).

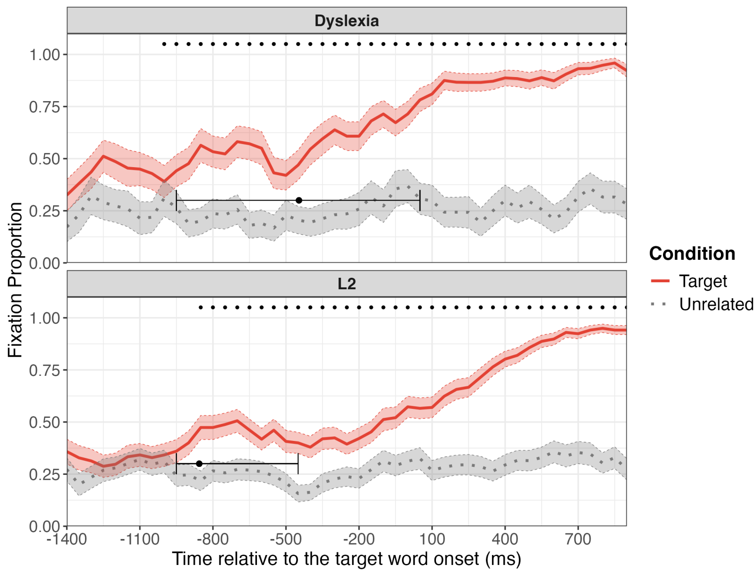


Figure 2.15: Divergence point analysis of mean fixation proportions over time for the target and unrelated conditions. Fixation proportions to the critical word are plotted over time relative to the target word onset (0 ms), participants with dyslexia and Tyrolean L1 speakers (L2).

Phonological competition effects in the anticipatory window. The model showed no significant effect of phonological competition across all groups. Difference smooth plots (Figure 2.9a-c) showed no significant time window where phonological competitor looks exceeded unrelated condition looks in any group. This finding suggests that participants did not show evidence of phonological form preactivation of highly predictable target words during anticipatory processing. This pattern suggests that anticipatory processing in highly predictable contexts may primarily involve semantic rather than phonological preactivation, or that phonological competition effects were not detectable under these experimental conditions.

To assess the sensitivity of the present design to detect phonological competitor effects, a simulation-based power analysis was conducted using linear mixed-effects models fitted to empirical logit-transformed fixation proportions (Barr, 2008). Variance components were estimated from the present data over the -1500 to -840 ms pre-target window – the interval identified by Ito (2024) as the period during which phonological competitor effects emerge in studies using a 2000 ms preview, as in the present experiment. Using this empirically grounded window avoids inflating power estimates by assuming a sustained effect across periods where the effect is known to be absent. Models were fitted separately for the full sample and for the L1 Italian typical reader group, as this group represents the population for which phonological prediction effects are most robustly documented in the literature and is therefore the most theoretically relevant subgroup for evaluating sensitivity. The most complex random effects structure that converged without singularity was retained in each case. The effect size was not derived from the present data but from the meta-analytic estimate reported by Ito (2024): a 5% absolute increase in fixation proportion to the phonological competitor relative to the unrelated baseline, corresponding to the peak effect size identified across studies using a target-absent design with a 2000 ms preview. One thousand datasets were simulated for each scenario and analysed with linear mixed-effects models with by-subject and by-item random intercepts and maximum likelihood estimation. The simplified random effects structure was justified by the negligible variance components observed. Simulations were validated by confirming Type I error rates close to the nominal 5% level under a null effect.

Estimated power was 22.8% for the full sample (n = 54) and 11.5% for the L1 Italian typical reader group (n = 23). This low sensitivity reflects both the relatively small per-group sample size and the high residual variance in the data, likely exacerbated by the lower temporal resolution of the eye-tracker (60 Hz), which limited the number of observations contributing to each condition cell. These estimates indicate that the absence of a phonological competitor effect should be interpreted with caution, as the design had limited sensitivity to detect effects of the magnitude reported in the literature.

2.2.6 Results. Resolution Window

A model comparison using `compareML` showed that the full model had a significantly better fit compared to the null model without the parametric interaction between condition and group (ΔfREML = –8,967.05, Δdf = 22.0, $p < .0001$, AIC difference = 18,159.10). Similarly, including the autocorrelation parameter (rho) substantially improved the model fit (ΔfREML = –6,287.76, Δdf = 0, $p < .0001$, AIC difference = 12,591.59). See Appendix D.2 for a full model summary.

In the resolution window, GAMM analysis revealed that the proportion of looks to the critical picture in the target condition was significantly higher than in the unrelated condition for all groups, with this difference emerging from the onset of the time window (Figures 2.16, 2.18a-c).

Group differences in word recognition. While all participant groups demonstrated preactivation of the target word before its onset (as evidenced in the prediction window analysis), the time course of lexical access revealed subtle but meaningful differences between groups. Smooth function comparisons indicated that Italian L1 typical readers maintained consistently

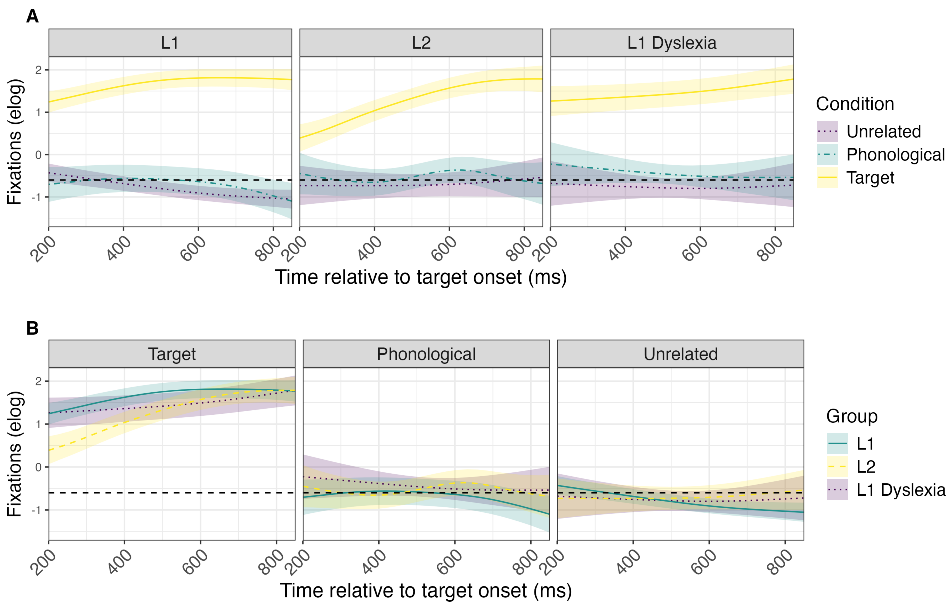
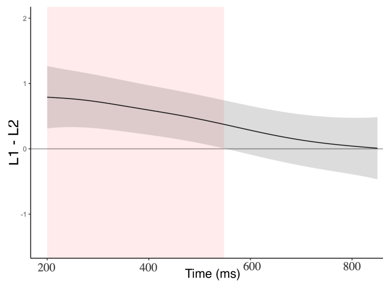


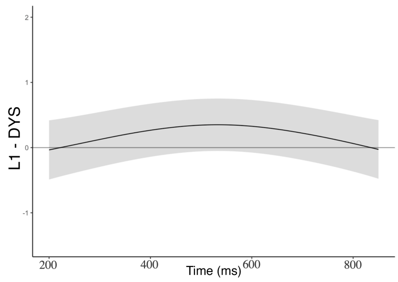
Figure 2.16: GAMM model estimates of looks to the critical object over time. (A) Comparison by condition for each group. (B) Comparison by group for each condition. The x-axis represents time relative to the target word onset (0 ms), and the y-axis shows the proportion of looks directed to the critical word transformed into empirical logit.

higher overall fixation proportions to the target compared to Tyrolean L1 speakers until approximately 548 ms post-onset (mean difference = 0.616), after which the fixation patterns converged (see Figure 2.17a). Similarly, L2 participants showed slightly lower looks to the target compared to participants with dyslexia during the early resolution window, with this difference persisting until approximately 331 ms post-onset (mean difference = 0.677) (Figure 2.17c). In contrast, typical L1 speakers did not differ from participants with dyslexia in the overall efficiency of lexical access, with both groups fixating on the target with similar amplitudes (Figure 2.17b).

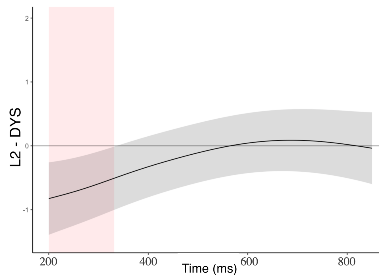
Phonological competition effects during lexical access. Regarding phonological competition, none of the groups showed significant activation of phonological competitors after target word onset, as indicated by confidence intervals that consistently included zero throughout the analysis window (Figures 2.18a-c). The non-significant effects of the phonological



(a) L1 - L2



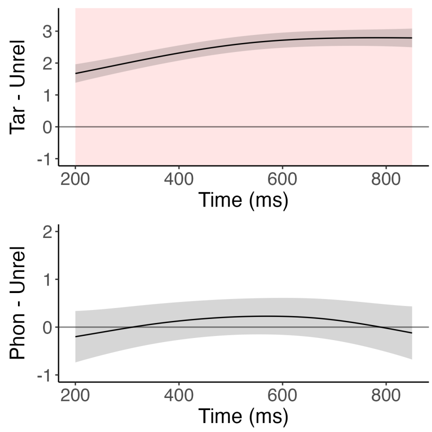
(b) L1 - Dyslexia



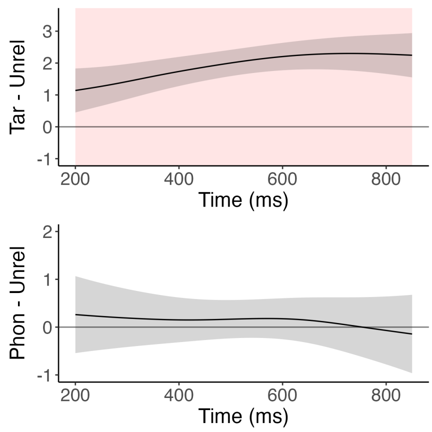
(c) L2 - Dyslexia

Figure 2.17: Panels show differences in empirical logit-transformed fixation proportions over time between groups in the *target* condition across the resolution time window for (a) L1 Italian vs. L1 Tyrolean speakers, (b) L1 Italian typical readers vs. L1 Italian participants with dyslexia, and (c) L1 Tyrolean speakers vs. L1 Italian participants with dyslexia. The x-axis represents time relative to target word onset (0 ms).

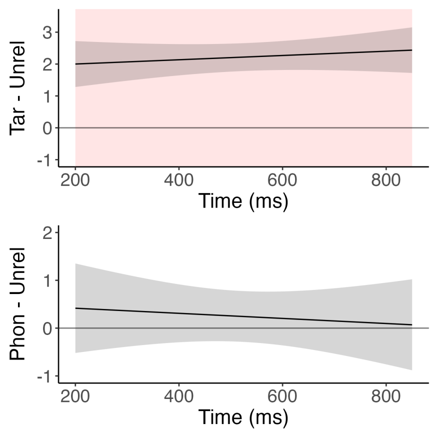
condition compared to unrelated items suggest that once the target word began unfolding, participants rapidly resolved any potential phonological competition in favor of the target. This efficient resolution of competition was consistent across all groups.



(a) L1 Italian typical readers



(b) L2 Italian L1 Tyrolean speakers



(c) Participants with dyslexia

Figure 2.18: Panels show differences in empirical logit-transformed fixation proportions over time between the target and unrelated conditions (upper graph in each panel) and the phonological and unrelated conditions (lower graph in each panel) for (A) L1 Italian typical readers, (B) L2 Italian L1 Tyrolean speakers, and (C) participants with dyslexia. The x-axis represents time relative to target word onset (0 ms).

2.3 Discussion

The integrated analysis across all participant groups revealed both similarities and differences in spoken language processing. Where differences emerged, they were primarily quantitative in nature, manifesting as differences in timing or magnitude of effects. All three groups showed anticipatory looks to the critical word in the target condition, confirming robust preactivation of upcoming linguistic input during spoken language comprehension. This finding aligns with the substantial body of literature that considers predictive processing to be a fundamental component of language comprehension (e.g., Altmann & Kamide, 1999; DeLong et al., 2005; Federmeier, 2007).

Crucially, the results revealed distinct patterns in the time course and magnitude of anticipatory processing between L1 and L2 speakers. Italian L1 typical readers demonstrated earlier and stronger anticipatory engagement relative to Tyrolean L1 speakers. In contrast, Tyrolean speakers during spoken comprehension in their L2 (Italian) demonstrated a more gradual increase in looks to the highly predictable critical word. Group differences in fixation proportions to the target emerged over time, reaching statistical significance approximately 970 ms before target onset. Moreover, Italian L1 speakers initiated anticipatory processing earlier than Tyrolean speakers, with a mean divergence point approximately 250 ms earlier. Although this difference should be interpreted cautiously, given overlapping confidence intervals and variability, the observed pattern aligns with previous findings in L2 comprehension showing preserved anticipatory processing with relatively small temporal delays that are often modulated by proficiency and language experience (Chambers & Cooke, 2009; Ito et al., 2018; Tagliani et al., 2025; Van Bergen & Flecken, 2017). Across studies, highly proficient L2 speakers often show anticipatory effects comparable to those of native speakers, whereas learners with lower or intermediate proficiency tend to show modest delays in onset latency. To assess whether such individual-difference factors contributed to the variability observed in the present data, I examined correlations between divergence point estimates and LEAP-Q measures within the Tyrolean group. As reported in Appendix G, none of these measures reliably predicted divergence point timing, suggesting that, in the present sample, differences in anticipatory onset among L2 speakers were not systematically related to self-reported proficiency or exposure. However, this null correlation should be interpreted cautiously, given the small Tyrolean subsample and possible limitations of self-report proficiency measures.

Several accounts may explain why L2 comprehenders in this study appeared less certain in their anticipatory fixations, despite showing broadly similar divergence points as L1 speakers. The weaker lexical-semantic links hypothesis receives indirect support from the present data. This account proposes that prediction in the L2 is explained by weaker links between lexical items' semantics and phonology, which results in slower lexical access and consequently slower predictions during language comprehension (Dijkgraaf et al., 2019). Similarly, vocabulary size has been shown to correlate with prediction efficiency, indicating that the strength of lexical representations is critical for anticipatory processing (Borovsky et al., 2012). However, recent evidence from Fernandez, Shehzad, and Hadley (2025) suggests that delays in semantic prediction due to weaker semantic network efficiency and the speed of lexical access may be age-specific. In their study comparing younger adult L1 and L2 speakers, Fernandez, Shehzad, and Hadley found that predictions based on semantic spreading activation were quite similar between groups, contrasting with the delays they observed in older adults (mean age 68.87 years). A second possible explanation emphasizes the non-selective nature of the bilingual lexicon (see Section 1.4.1). If lexical representations in both languages are activated in parallel, they may compete for selection and create slower or less robust input for downstream processing (Hopp, 2018; Ivanova & Costa, 2008). Non-selective activation can make lexical retrieval more diffuse, thereby reducing the efficiency of predictive mechanisms. Finally, the processing resources hypothesis appears less applicable to the present data. This account highlights that L2 comprehenders may rely on qualitatively similar mechanisms as L1 comprehenders but with reduced efficiency due to weaker syntactic processing (Clahsen & Felser, 2006), weaker semantic networks (Ivanova & Costa, 2008), and less automatic comprehension (Segalowitz & Hulstijn, 2009). Prediction in L2 may therefore place higher demands on domain-general executive resources such as memory and cognitive control (Ryskin et al., 2020). However, the present experimental design featured relatively simple sentences with highly constraining verb-based predictions, placing minimal demands on syntactic processing or working memory. The current data cannot definitively isolate the relative contribution of each mechanism. However, these explanations are not mutually exclusive: weaker lexical–semantic links, non-selective cross-linguistic activation, and higher processing demands may jointly contribute to why prediction in the L2 often appears less strong.

The comparison between L1 typical readers and participants with dyslexia did not reveal any significant differences in the temporal dynamics of anticipatory eye movements, as indicated by both the GAMM and divergence point analyses. Importantly, a wide confidence interval for the mean divergence point between the two groups (150 – 1300 ms), mainly influenced by the dyslexia group, suggests greater inter-individual variability in predictive processing within this experimental group compared to typical readers. Several factors may contribute to this heterogeneity. First, the relatively small sample size of participants with dyslexia may have contributed to the observed variability in the data. The heterogeneity might also arise from differences in cognitive profiles among individuals with dyslexia, a phenomenon that has been well-documented in previous research (Heim & Grande, 2012; Pacheco et al., 2014; Pennington et al., 2012). For instance, Pacheco et al. (2014) identified two distinct clusters of dyslexic children based on phonological awareness, rapid naming, and other cognitive abilities. This aligns with Heim and Grande's (2012) concept of dyslexia "fingerprints," referring to distinct cognitive profiles of individuals with dyslexia and their corresponding neurobiological patterns. These findings underscore the importance of considering individual differences when studying dyslexia. That is why in my subsequent experiment (discussed in Chapter 4), I included additional measures, such as the Rapid Automatized Naming (RAN) Task (CTTOP), Stroop Task, Adult Reading History Questionnaire (ARHQ), Peabody Picture Vocabulary Test (PPVT), and Author Recognition Test (ART), for more tailored comparisons. This represents a methodological refinement based on insights gained from Experiment 1, where the heterogeneity in dyslexic participants' responses highlighted the need for more detailed cognitive profiling.

Interestingly, the comparison between L2 speakers and L1 speakers with dyslexia revealed no significant differences in anticipatory processing. Thus, despite different underlying factors (second language processing vs. reading disorder), these groups showed similar patterns in both the magnitude and time course of their anticipatory looks to the target.

In the resolution time window, all groups maintained a strong preference for the target after its onset compared to unrelated items, indicating that predictions were successfully confirmed and integrated following auditory input. However, L2 speakers maintained a lower proportion of looks to the target word compared to L1 speakers until approximately 550 ms post-onset, after which the fixation patterns converged. This suggests that despite a slight temporal delay, Tyrolean speakers eventually achieved similar levels of target identification. One possibility is that, once sufficient bottom-up information accumulates, the lexical input overrides earlier delays caused by weaker lexical–semantic links, slower access, or higher processing demands in the L2. In other words, predictive processing may be less efficient in L2, but recognition processes allow listeners to "catch up" with L1 speakers in the resolution window. In the present experiment, all sentences were predictable, so it was not possible to directly assess whether L2 and L1 groups differed in lexical recognition in the absence of prediction. This issue is addressed in the subsequent English experiment, which included unpredictable sentences and thus provides an additional test of group comparability in lexical identification (see Chapter 4). Similarly to the anticipatory window, participants with dyslexia did not show any differences in looks to the target compared to typical readers. Finally, the comparison between L2 speakers and L1 speakers with dyslexia revealed only a slight difference in lexical access, with L2 speakers showing a lower proportion of looks to the target in the earliest part of the post-onset fixation window (roughly the first 130 ms of the observable fixation response, corresponding to $\sim$ 330 ms after the acoustic target onset once saccadic latency is taken into account).

The absence of a significant phonological effect directly addresses my first research question regarding whether phonological features of upcoming predictable words are retrieved before the acoustic onset. The results do not provide evidence for phonological form preactivation, as participants did not show increased looks toward phonological competitors compared to unrelated items in the anticipatory window across any group. This pattern diverges from studies that have found phonological form preactivation (DeLong et al., 2005; Ito et al., 2018, 2020; Li & Qu, 2023; Li et al., 2022) but aligns with those suggesting that prediction operates mainly at semantic or conceptual levels without extending to detailed phonological representations (Ito & Husband, 2017; Ito et al., 2017b; Nieuwland et al., 2018). One possible explanation is that the strong contextual constraints in the sentence stimuli allowed listeners to rapidly discard phonologically similar but semantically incongruent competitors. In addition, onset phoneme overlap between targets and competitors could have been too small, leading to no observable effect (Ito, 2024). A further methodological consideration is statistical power. A simulation-based power analysis, grounded in the variance components estimated from the present data and the meta-analytic effect size reported by Ito (2024), indicated that this study had an estimated power of 22.8% for the full sample and only 11.5% for the L1 Italian typical reader group. These low estimates reflect two compounding factors: the small per-group sample size and the higher residual variance attributable in part to the lower temporal resolution of the Gazepoint GP3 HD eye-tracker (60 Hz). The absence of a phonological effect is therefore consistent with the design being underpowered rather than, or in addition to, reflecting a genuine absence of phonological preactivation. A more sensitive test is provided by Experiment 2 (see Chapter 4), which employed a higher-precision tracker (EyeLink 1000, 500 Hz) and a larger participant sample.

Furthermore, I found no evidence of phonological form activation in the resolution window. It is typically expected that as the word unfolds, cohort competitors sharing initial phonological features with the target would be briefly activated (Allopenna et al., 1998; Huettig & McQueen, 2007). However, the absence of such effects in the present study may reflect prediction-induced inhibition of competitors, i.e., high cloze probability of the sentences and the nature of the given task itself – to look at the mentioned object – likely induced participants to commit strongly to the predicted target, preempting any activation of phonological form competitors. Haeuser and Borovsky (2024) demonstrated that when words are strongly predicted in constraining contexts, the predicted target actively suppresses phonologically similar competitors during word recognition. This suppression mechanism, consistent with competitive word recognition models (McClelland & Rumelhart, 1981), may explain why phonological competitors remained inactive even after target word onset in highly predictive sentences. These findings contrast with those by Ito et al. (2018), who reported that L2 participants activated phonological form information associated with the target word after its onset via priming, while L1 speakers did not (see Section 1.4.1). One potentially relevant difference between the two studies concerns the age of first exposure to the L2. In the present study, Tyrolean speakers reported first exposure to Italian at a relatively early age (M = 5.8 years, SD = 1.8, range = 3–10 years), whereas the L2 participants tested by Ito et al. (2018) reported a later mean age of first exposure to English (M = 10 years, range = 5–15 years). Earlier exposure could plausibly lead to tighter coupling between contextual–semantic constraints and phonological processing, such that activation remains focused on the contextually licensed target and does not spread to phonologically similar but contextually irrelevant competitors, thereby reducing the emergence of post-onset phonological priming effects. To further investigate whether post-onset phonological competitor effects depend on sentence predictability or instead reflect more general activation dynamics in L2 comprehension, I included an unpredictable condition in the second study (discussed in Chapter 4). This manipulation allows for a more direct test of whether phonological activation in the resolution window is modulated by contextual constraint.

In sum, this integrated analysis provides evidence for both differences and similarities in anticipatory language processing across different population groups. All groups engaged in predictive processing, but with distinct temporal dynamics: typical Italian readers demonstrated earlier and more robust anticipatory looks to target objects before hearing the corresponding words compared to L2 speakers. The observed temporal differences contribute to ongoing debates about whether prediction mechanisms operate fundamentally differently in L1 versus L2 processing or whether the same mechanisms operate with different efficiency levels. The findings, which show robust anticipatory processing with only slight delays in the Tyrolean group, provide support for the shared mechanisms hypothesis. The temporal delay and reduced magnitude of anticipatory looks in L2 participants suggest that these predictive mechanisms operate with reduced efficiency or under greater resource constraints in this population group. This pattern aligns with Kaan's (2014) proposal that L2 prediction differs from L1 prediction primarily in degree rather than kind. Observing convergence in the resolution window further supports this interpretation – once sufficient bottom-up input becomes available, the L1-L2 difference disappears, suggesting that underlying language comprehension mechanisms are fundamentally similar. Regarding the dyslexia group, the results provide evidence for intact predictive processing mechanisms in individuals with dyslexia. Despite well-documented difficulties with phonological processing and reading, the current findings reveal that their ability to generate context-based predictions is comparable to that of typical readers. Moreover, both experimental groups showed comparable results in key measures of anticipatory processing. The phonological null result should be interpreted in light of the limited statistical power of the present design, and a more definitive test is reserved for Experiment 2. Nonetheless, to the extent that the null result is meaningful, it may suggest that strong predictive mechanisms do not necessarily extend to phonological form activation during anticipatory processing. Even when semantic prediction is robust, phonological processing may either fail to extend to the level of phonological form or remain tightly constrained by contextual expectations, such that activation is focused on the predicted target and does not spread to phonologically similar but contextually irrelevant alternatives. In this view, strong contextual constraints may alter the dynamics of lexical access, as reflected in the absence of phonological competitor effects during the word recognition phase, potentially allowing comprehenders to bypass typical phonological competition in favor of confirming a highly predicted target.

Chapter 3

Experiment 1b: Grammatical Gender as a Predictive Cue in Italian

The second phase of the Italian experiment investigated how participants with dyslexia and L2 speakers utilize grammatical gender cues in predictive language processing. While the goal of the first part of the experiment (Experiment 1a, discussed in Chapter 2) was to examine semantic and phonological prediction mechanisms, this part focused specifically on morphosyntactic prediction through gender agreement systems. The study addressed three primary research questions:

1. Do individuals use grammatical gender cues (e.g., determiners) in prediction, and are there differences between participants with dyslexia and L2 speakers compared to L1 typical readers?

2. Does cross-linguistic gender congruency between Italian and Tyrolean modulate predictive processing in bilingual participants? Specifically, do Tyrolean L1 speakers show enhanced prediction when grammatical gender of the target overlaps between the two languages?

3. Does the presence of cognates[1] among competitor items influence prediction in L2 participants?

3.1 Method

3.1.1 Participants

The same 57 participants as in Experiment 1a participated in this part, maintaining the group division (see Section 2.1.1). For the present analyses, only two participants were excluded during data cleaning, compared to three exclusions in Experiment 1a, as one Tyrolean participant met the quality criteria in this part but not in the previous one (see Section 3.2 for details on the data cleaning procedure). As a result, the final sample for this part consisted of 55 participants (see Table 3.1).

Table 3.1: *Demographic characteristics of study participants by group (Experiment 1b)*

	Italian L1		Tyrolean L1
Characteristic	Typical	Dyslexia	
Sample size (n)	23	12	20
Gender (f/m)	14/9	8/4	12/8
Age (years)	23.1 (4.7)	25.6 (8.5)	25.95 (8.7)
Education level (% secondary/higher)	87/13	75/25	74/26

Note. Values show Mean (Standard Deviation).

3.1.2 Materials

The experimental task involved a visual world paradigm where participants viewed arrays of four pictures (one target and three distractors) while listening to Italian sentences beginning with "*Guarda la loro/il loro...*" [Look at their...]. Participants were introduced to the Rossi family, who enjoy collecting various items from objects to animals, and were instructed to look at the mentioned object on the screen as quickly as possible (see Figure 3.1). All sentence stimuli were recorded by the same female native

[1]For more information on cognates, please refer to Section 1.4.1 of Chapter 1.

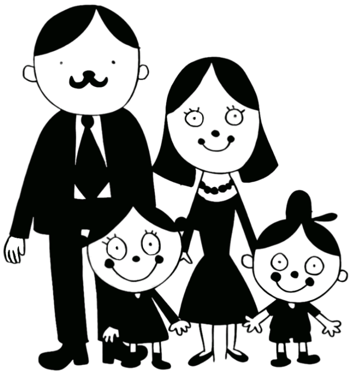


Figure 3.1: The Rossi family introduction screen shown to participants before the experimental task.

speaker of Italian at the same time as the stimuli used in Experiment 1a (see Section 2.1.2.2). Because all sentences followed an identical syntactic structure and differed only in the sentence-final noun, the recordings were subsequently segmented using audio-editing software (Audacity 3.7.1; Audacity Team, 2024) so that the same sentence onset (*Guarda il loro* / *Guarda la loro*) was used across all items of the corresponding gender and combined with the remainder of each sentence. This prevented item-specific coarticulatory cues between the determiner and the noun, allowing eventual anticipatory effects to be attributed to grammatical gender marking on the determiner. The possessive pronoun *loro* [their] was included in all sentences to extend the time window between the article and the noun, providing participants with additional processing time for eventual anticipatory looks.

The experiment followed a 3-condition within-subjects design (prediction congruent, prediction incongruent, non-predictive baseline). A total of 18 critical items were constructed, with 6 items per condition (see Appendix B). Each item appeared in only one condition, and all participants saw all items presented in a randomized order. In addition, 3 practice trials were included at the beginning of the experiment. Each trial displayed four pictures (one target and three distractors). No filler trials were included.

3.1.2.1 Experimental Conditions.

Three main experimental conditions were implemented to investigate grammatical gender-based prediction (see Figure 3.2 for an example display).

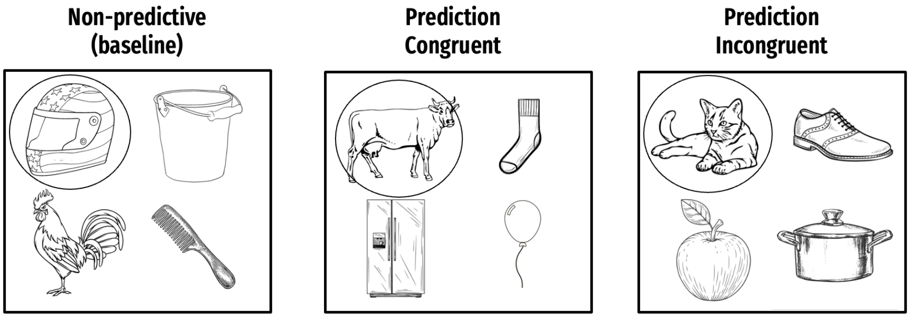


Figure 3.2: Experimental conditions for sentences: *Guarda il loro casco* [Look at their helmet] for the non-predictive condition; *Guarda la loro mucca* [Look at their cow] for the prediction congruent condition; *Guarda il loro gatto* [Look at their cat] for the prediction incongruent condition. Targets are encircled.

The three experimental conditions were the following:

- **Non-predictive Condition (baseline)**: All four pictures depicted objects of the same grammatical gender in both Italian and Tyrolean, eliminating the possibility of gender-based prediction. For example, all objects were masculine in both languages: *il casco* (M)[2] / *der Helm* (M) [helmet], *il secchio* (M) / *der Kiebl* (M) [bucket], *il pettine* (M) / *der Kompl* (M) [comb], and *il gallo* (M) / *der Gigger* (M) [rooster] (see Figure 3.2). Thus, no anticipatory eye movements toward the target were expected in this condition.
- **Prediction Congruent**: One object (the target) differed in grammatical gender from the other three. The grammatical gender of depicted objects was consistent across both languages. For instance, in the display for *Guarda la loro mucca* [Look at their cow], the target *la mucca* (F) / *die Kuah* (F) appeared together with three masculine distractors: *il palloncino* (M) / *der Luftballon* (M) [balloon], *il calzino*

[2]Morphological labels follow the Leipzig Glossing Rules (Comrie et al., 2015) using F = feminine, M = masculine.

(M) / *der Sock* (M) [sock], and *il frigorifero* (M) / *der Eiskoschtn* (M) [fridge]. In this configuration, the determiner uniquely pointed to the target in both languages and should therefore support anticipatory looks.

- **Prediction Incongruent**: The target's grammatical gender diverged across the two languages, such that the Italian article supported prediction of one picture while Tyrolean gender pointed to a different one. In the display for *Guarda il loro gatto* [Look at their cat], the Italian target *il gatto* (M) contrasted with the Tyrolean *die Kotz* (F). The three distractors showed the opposite pattern, being feminine in Italian but masculine in Tyrolean: *la scarpa* (F) / *der Schuah* (M) [shoe], *la mela* (F) / *der Epfl* (M) [apple], and *la pentola* (F) / *der Topf* (M) [pot]. Thus, gender-based prediction was congruent within each language but mutually incompatible across Italian and Tyrolean.

Additionally, to examine the influence of cross-linguistic lexical similarity on predictive processing in bilingual participants, a subset of the experimental items included Italian–Tyrolean cognate pairs. Cognates were distributed across conditions such that 8 of the 18 experimental trials contained both a target and one competitor that were Italian–Tyrolean cognates, resulting in two of the four pictured objects per display being cognates (see Figure 3.3). Including competitor cognates, rather than only target cognates, served to avoid systematic associations between cognate status and target position. In total, 22 cognate words were included across the experiment:18 in experimental trials and 4 in practice trials.

Cognate status was expected to interact with the prediction manipulation depending on condition. Target cognates may facilitate recognition in congruent conditions, whereas cross-linguistic similarity may have competing effects in incongruent trials (see Section 3.1.4). As discussed in Section 2.1.1, L1 participants grew up with Italian as their only home language and reported no knowledge of Tyrolean. While all had experience with additional languages (most commonly English), these were acquired later in life and not from birth, so they cannot be considered simultaneous

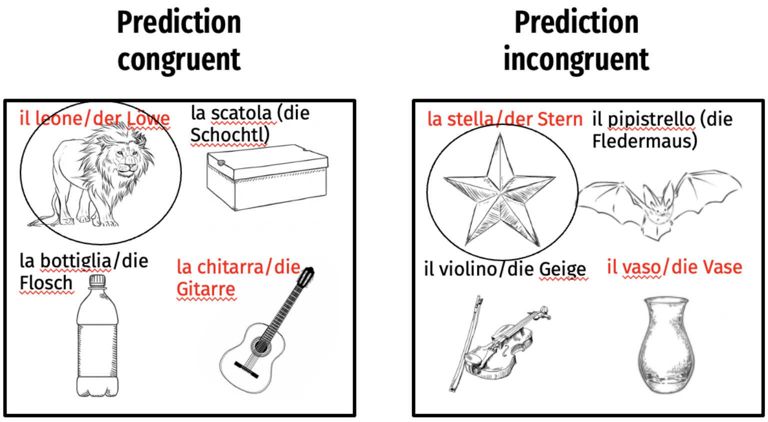


Figure 3.3: Examples of prediction *congruent* and *incongruent* trials illustrating the cognate manipulation. Items shown in red are Italian–Tyrolean cognates. In the congruent example (left), the target *il leone/der Löwe* ('the lion') and the competitor *la chitarra/die Gitarre* ('the guitar') are cognates, while the other distractors (*la bottiglia/die Flosch* 'the bottle', *la scatola/die Schochtl* 'the box') are non-cognates. In the incongruent example (right), the target *la stella/der Stern* ('the star') and the competitor *il vaso/die Vase* ('the vase') form the cognate pair, with non-cognate distractors (*il pipistrello/die Fledermaus* 'the bat', *il violino/die Geige* 'the violin') completing the display.

bilinguals. Accordingly, the Italian–Tyrolean congruency and cognate factors are theoretically targeted at the Tyrolean group, while any cognate facilitation in Italian L1 (e.g., via English) is incidental and not expected to interact with the congruency manipulation.

3.1.3 Procedure

Experiment 1b followed immediately after Experiment 1a and began with three practice trials. Whereas some sentences in Experiment 1a did not refer to any of the displayed objects, each sentence in this phase corresponded to one of the picture objects. Participants were introduced to the Rossi family, who enjoy collecting various items, from objects to animals (Figure 3.1). They were informed that they would now hear what the Rossi family collects, and their task was to look at the mentioned object on the screen as quickly as possible.

After the practice trials, participants were shown four pictures (one target and three distractors) (see Figure 3.2) while listening to sentences that began with "*Guarda la loro/il loro...*" [Look at their...]. As in the previous part (Experiment 1a), the visual presentation of the sentence stimuli began 2000 ms before the acoustic onset of the critical word. However, since sentences in this part were much shorter compared to the first part, pictures appeared on the screen before the sentence started (see Figure 3.4). No feedback was given during the experiment.

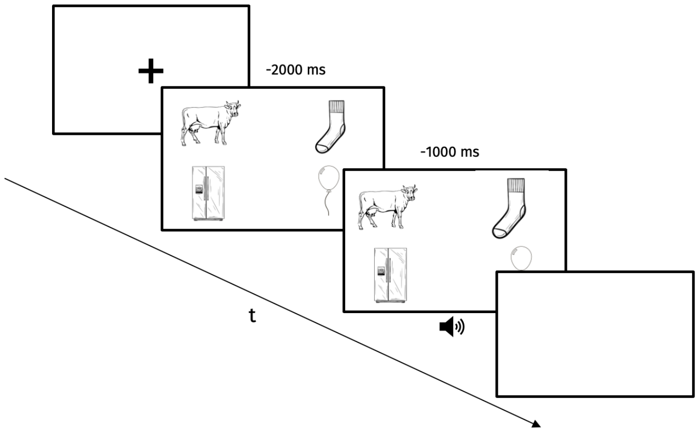


Figure 3.4: Trial procedure and timing for Experiment 1b.

3.1.4 Predictions

Based on the theoretical framework and empirical findings reviewed in Chapter 1, I formulated the following predictions for the present study.

Grammatical gender as a predictive cue. Research using both ERP and eye-tracking paradigms has demonstrated that listeners can exploit grammatical gender marking on determiners to anticipate upcoming nouns in real time (see Section 1.3.2). In visual-world studies, native speakers of languages with grammatical gender systems show anticipatory fixations to gender-matching referents as soon as the article is heard, before the noun itself is spoken (Bosch & Foppolo, 2022; Huettig & Brouwer, 2015). Given this evidence, I predicted that Italian L1 typical readers would show strong anticipatory looks to gender-congruent targets in both prediction congruent and prediction incongruent conditions, reflecting efficient exploitation of grammatical gender cues during real-time comprehension. For Tyrolean

L1 speakers (L2 Italian), the literature on morphosyntactic prediction in L2 comprehension indicates that while L2 speakers can use gender cues predictively, such effects are typically weaker, delayed, or more variable than in L1 comprehension, and are often modulated by cross-linguistic overlap and proficiency (Dussias et al., 2013; Hopp, 2013; Hopp & Lemmerth, 2018). Because both Italian and Tyrolean encode grammatical gender on determiners, this structural similarity should facilitate gender-based prediction. However, differences in cue reliability, lexical gender knowledge, and overall L2 proficiency may result in reduced or delayed anticipatory effects compared to native Italian speakers. For Italian L1 speakers with dyslexia, Huettig and Brouwer (2015) demonstrated that adults with dyslexia show delayed anticipatory eye movements when using gender cues to predict upcoming referents (see Section 1.4.2). Importantly, their findings indicate that predictive use of grammatical gender is not absent, but rather less efficient, in dyslexia. Accordingly, I expected Italian L1 participants with dyslexia to show a preserved ability to use gender cues, albeit with delayed temporal dynamics compared to typical readers.

Cross-linguistic gender congruency effects. As reviewed in Section 1.4.1, cross-linguistic similarity at the lexical and morphosyntactic levels might determine whether L2 comprehenders can exploit grammatical gender cues predictively. A *gender congruency effect* has been observed in both adult L2 learners and bilingual children: when a noun carries incongruent gender values across the two languages, predictive processing is delayed (Bosch & Foppolo, 2022; Hopp & Lemmerth, 2018). These findings indicate that cross-language gender congruency can modulate the efficiency of morphosyntactic prediction. In the present study, Tyrolean L1 speakers should therefore show faster and stronger anticipatory effects when the Italian target gender matches Tyrolean gender (congruent condition) compared to when the genders differ across languages (incongruent condition). Incongruent gender is expected to create interference through competing activations from the two language systems. Italian L1 participants, by contrast, should show no difference between cross-linguistically congruent and incongruent conditions, as they do not know Tyrolean and therefore should not experience cross-linguistic interference.

Cognate effects on gender prediction. Research on cognates demonstrates that lexical items sharing form and meaning across languages produce facilitation effects during comprehension, reflecting parallel activation of both language systems even in monolingual contexts (Costa et al., 2000; De Groot & Keijzer, 2000, see Section 1.4.1). Critically, however, the direction of cognate effects can vary depending on whether form overlap supports or conflicts with target identification. When grammatical gender aligns across languages (congruent trials), cognate targets should benefit from convergent cues: both the Italian gender-marked determiner and cross-linguistic phonological similarity support rapid identification. Cognate competitors in such trials, by contrast, should produce minimal interference because the gender cue unambiguously favors the target. For example, in a trial where the target is *il leone* (M) – *der Löwe* (M) [lion], a cognate competitor such as *la chitarra* (F) – *die Gitarre* (F) [guitar] should not strongly compete, as the masculine determiner clearly excludes it (see Figure 3.3). In cross-linguistically incongruent trials, however, cognate competitors may increase interference when gender cues diverge between the two languages. For instance, in a trial with target *la stella* (F) – *der Stern* (M) [star], the Italian feminine determiner supports the target in Italian, but a cognate competitor such as *il vaso* (M) – *die Vase* (F) [vase] receives additional activation from Tyrolean, where it is feminine. This cross-linguistic mismatch should increase competition and potentially slow or reduce anticipatory looks to the Italian target (see Figure 3.3). For Italian L1 speakers, no interaction with Italian–Tyrolean gender congruency is predicted. Any general cognate effect (e.g., due to overlap with other known languages such as English) would be expected to manifest as a main effect of cognate status rather than an interaction with the cross-linguistic gender congruency manipulation.

3.2 Data Analysis and Results

3.2.1 Data Cleaning and Analysis

Gaze data from Experiment 1b were preprocessed following the same procedure as in Experiment 1a (see Section 2.2.3), using the eyetrackingR package (Dink & Ferguson, 2015). Fixation proportions were calculated for each 50 ms time bin, and trials with excessive trackloss (i.e., > 27%) were excluded, resulting in the removal of 76 trials (6% of the dataset). As before, trials in which the participant never fixated on the target object throughout the entire trial were excluded. Additionally, participants whose usable trial count fell more than 1.5 standard deviations below the group mean (i.e., fewer than 10 trials) were excluded, resulting in the removal of two participants: one from the L1 typical reader group and one from the dyslexia group. Updated participant characteristics for Experiment 1b are reported in Table 3.2, which includes the recalculated means for demographics and language background (LEAP-Q) measures. These values differ only minimally from those reported in Experiment 1a (Section 2.1, Table 2.2), reflecting the inclusion of one additional Tyrolean participant whose data met quality criteria here.

The analysis employed Generalized Additive Mixed Models (GAMMs) to examine fixation patterns across two critical time windows using the `mgcv` (Wood, 2017) and `itsadug` (Van Rij et al., 2015) packages in R. The anticipatory window (-250 ms to 200 ms relative to target word onset) captured the period from determiner onset to the acoustic onset of the target word, allowing to investigate predictive processing. The resolution window (200 ms to 900 ms relative to target word onset) examined the period in which lexical resolution and integration occur.

All models predicted the empirical logit-transformed proportion of looks to the target object (*elog*). The core structure of the models included:

- A smooth term over time `s(time)` to model the non-linear trajectory

Table 3.2: *Language background characteristics from LEAP-Q assessment (Experiment 1b)*

Characteristic	Italian L1		Tyrolean L1
	Typical	Dyslexia	
L1 Language Experience:			
Daily exposure to L1 (%)	67.4 (17.5)	66.0 (19.6)	66.9 (19.6)
L2 (Italian) Language Experience:			
Age of acquisition (years)	–	–	5.9 (1.8)
Daily exposure (%)	–	–	20.1 (10)
Italian Academic/Professional Exposure:			
None	–	–	33%
1–5 years	–	–	17%
More than 5 years	–	–	50%
Self-Rated L1 Proficiency:			
Speaking	5.9 (0.4)	5.7 (0.7)	5.7 (0.7)
Listening	5.9 (0.4)	5.7 (0.6)	5.7 (0.7)
Reading	5.8 (0.5)	5.6 (0.9)	5.5 (0.9)
Writing	5.8 (0.6)	5.4 (1.0)	5.3 (1.0)
Average	5.9 (0.4)	5.6 (0.7)	5.6 (0.8)
Self-Rated L2 (Italian) Proficiency:			
Speaking	–	–	4.1 (1.1)
Listening	–	–	4.65 (0.7)
Reading	–	–	4.4 (0.9)
Writing	–	–	3.8 (1.2)
Average	–	–	4.2 (0.96)

Note. Values show Mean (Standard Deviation) unless otherwise specified. L1 = first language; L2 = second language. Italian Academic/Professional Exposure indicates the duration of time spent in schools and/or workplaces where Italian was the primary language of communication. Self-rated proficiency measured on a scale from 1 (lowest) to 6 (highest).

of looks.

- Fixed effects encoding experimental condition and participant group, implemented as a condition-by-group interaction `cg` with condition-specific smooths over time `s(time, by = condition)`.
- Random effects to account for variability across participants and items. Participant-level random smooths over time were modeled using factor smooths (`fs`; `s(time, participant, bs = ‘‘fs’’)`). Item-level random effects were modeled using random-effect smooths (`bs = ‘‘re’’`), since items did not repeat across conditions.

To address autocorrelation in the eye-tracking time series, the `rho` parameter was included in the resolution window models using `AR.start` argument. In the anticipatory window, *rho* was not included, given that the null models were generally preferred in this window, indicating limited condition effects. On the main models, significant interactions were further examined with pairwise smooth comparisons using `itsadug` functions and cluster-based permutation tests, which identify time windows of reliable divergence between trajectories. This procedure was applied consistently across all between- and within-group comparisons. Divergence point analysis was also used to estimate the onset of condition differences.

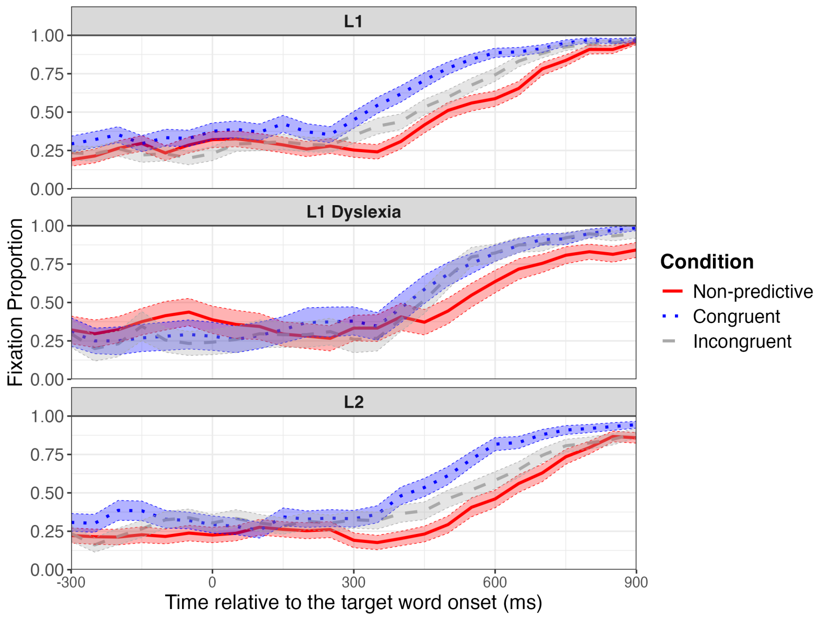


Figure 3.5: Mean fixation proportions on the critical word in congruent, incongruent, and non-predictive (baseline) conditions in the L1 (top graph), Dyslexia (middle graph), and L2 (bottom graph) groups. Time 0 ms represents the acoustic onset of the target word. Error bands indicate standard errors.

To visualize the data, I plotted time-course graphs of the proportion of looks to the critical word in each condition across all the participant groups (Figure 3.5).

3.2.2 Results. Anticipatory Window

The analysis of the anticipatory time window did not reveal any evidence of grammatical gender-based prediction across any participant group or condition. When comparing a full model including condition-group interactions against its corresponding null model without such interactions, adding the interaction term did not improve model fit (ΔfREML = -4.27, Δdf = 24, AIC difference = 10.13). To explore further, I collapsed the two predictive conditions (congruent and incongruent) into a single 'predictive' level [3] to see whether this might reveal effects (see Figure 3.6 for time-course graphs). Model comparison again showed that including condition–group interactions did not improve model fit (ΔfREML = –3.60, Δdf = 15, AIC difference =7.30). These findings suggest that participants did not engage in predictive processing of upcoming nouns based on determiner gender information during the pre-target window.

To examine whether cognate status modulated predictive processing, an additional model was fit, including a cognate × condition–group interaction, and compared to its corresponding baseline model without such interaction. This model comparison tested whether items including words with cross-linguistic similarity (cognates) modulated language processing in bilingual participants. Model comparison favored the simpler baseline model (ΔfREML = –7430.03, Δdf = 9, AIC difference = 14629.54, $p < 0.001$), indicating that adding cognate status substantially worsened model fit. These results suggest that cross-linguistic similarity between Italian and Tyrolean words did not modulate anticipatory eye-movement behavior in any group.

[3] For clarity, throughout the text I use the term *predictive condition* specifically for trials in which the determiner uniquely identified a single referent, such that grammatical gender could be used as a predictive cue. In the non-predictive condition, the same gender marking was shared by multiple referents, and thus the determiner did not allow unique anticipation.

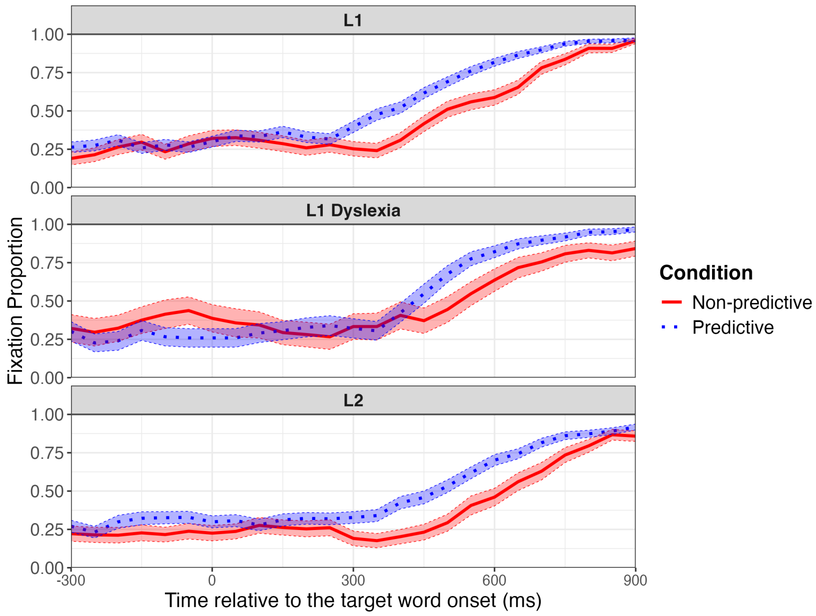


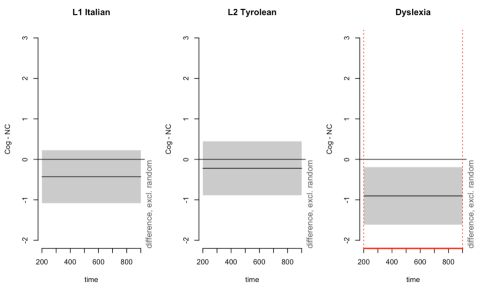
Figure 3.6: Mean fixation proportions on the critical word in the combined predictive (congruent + incongruent) and non-predictive (baseline) conditions in the L1 (top graph), Dyslexia (middle graph), and L2 (bottom graph) groups. Time 0 ms represents the acoustic onset of the target word. Error bands indicate standard errors.

3.2.3 Results. Resolution Window

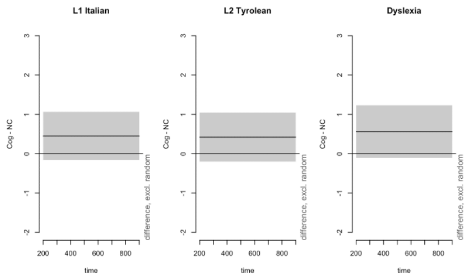
In contrast to the anticipatory window, the resolution window analysis revealed substantial condition and group effects. The full model with three-condition and group interactions provided a markedly better fit than the null model (ΔfREML = –11223.52, Δdf = 8, AIC difference = 22294.62, $p < .001$). This statistically significant difference in the models indicates that post-target processing effects vary across conditions and participant groups. To investigate whether cross-linguistic lexical overlap modulated target identification after word onset, I compared a model including the cognate × condition–group interaction to a baseline model without this term. The addition of the interaction significantly improved model fit in the resolution window (ΔfREML = –20.61, Δdf = 9, AIC difference = –41.16, $p < .001$), suggesting a potential role for cross-linguistic similarity in post-target processing. To further examine whether the cognate effect was selectively modulated by condition and group, I followed up with pairwise contrasts using estimated marginal means (Lenth & Piaskowski, 2025). The only statistically significant difference emerged in the dyslexia group in the baseline (non-predictive) condition, where non-cognate targets elicited more fixations than cognates (estimate = 0.90, SE = 0.36, p = 0.01) in the resolution time window. To visualize this effect, I plotted the difference in looks to the target between cognate and non-cognate items across all groups for the baseline condition (Figure 3.7a). However, no significant cognate effects were observed in any predictive condition or in other groups (all p > 0.1), including the L2 group where such modulation was hypothesised (Figure 3.7b-c). Given the small sample size in the dyslexia group and high variability, the effect is likely spurious, reflecting a Type I error due to sampling variability.

Moreover, given that the primary theoretical motivation for including cognate status was to test for facilitation in L2 speakers, I conducted a follow-up analysis restricted to the Tyrolean group. A model with a cognate × condition interaction was compared to a model without this term. Model comparison indicated that adding the interaction did not improve fit (ΔfREML = 1.63, Δdf = 3, p = .35, AIC difference = 0.07). Furthermore, pairwise comparisons of target fixation proportions revealed no significant differences between cognate and non-cognate items in any condition (all $p > .35$). These findings suggest that, contrary to predictions, cognate status did not modulate lexical resolution in the L2 group.

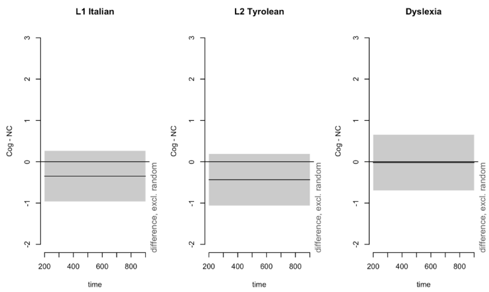
I next examined whether cross-linguistic gender congruency between Italian and Tyrolean modulated target identification. As outlined in the predictions, L2 speakers were expected to show faster processing in the congruent condition (where gender marking aligned across languages) compared to the incongruent condition (where gender marking conflicted). To assess these condition differences, statistical significance was evaluated using 95% confidence intervals around the GAMM smooth difference curves, computed with the `plot_difference()` function from the `itsadug` package (Van Rij et al., 2015) (see Appendix E.1 for the full model summary). Time periods where confidence intervals did not include zero were considered



(a) Non-predictive (baseline)



(b) Predictive congruent



(c) Predictive incongruent

Figure 3.7: Difference in looks to the target between items containing cognate versus non-cognate words across experimental conditions. Each panel displays the estimated time course of the difference in log-odds of target fixations (Cognate - Non-Cognate) for each participant group, with 95% confidence intervals. Positive values indicate greater fixations to cognate items, negative values indicate greater fixations to non-cognate items. (a) Non-predictive baseline condition (b) Predictive congruent condition (c) Predictive incongruent condition.

statistically significant. However, the analysis revealed no such differential effects. Critically, for the Tyrolean group, confidence intervals consistently included zero throughout the analysis window, indicating no significant differences between congruent and incongruent predictive conditions (see Figure 3.8). This pattern was consistent across all groups, with no significant differences observed between congruent and incongruent conditions in either the Italian L1 or dyslexia groups. This finding supported the decision to combine these conditions into a single *'predictive'* category for subsequent analyses, as cross-linguistic gender congruency did not differentially affect L2 processing in this paradigm.

Figure 3.8: Estimated difference in looks to the target over time between *congruent* and *incongruent* conditions in the L2 group (Tyrolean L1 speakers). The x-axis represents time relative to target word onset (in ms), and the y-axis shows the difference in empirical logit-transformed fixation proportions (*congruent* − *incongruent*). The solid line represents the estimated mean difference, with positive values indicating more target fixations in the congruent condition. The shaded band represents the 95% confidence interval across the resolution time window.

3.2.3.1 Combined Predictive vs. Non-Predictive Analysis: Within-Group Effects and Divergence Timing

Given the absence of congruency effects in the L2 group, I proceeded with a simplified model contrasting combining predictive conditions against a non-predictive control condition. This model incorporated autocorrelation correction (*rho* parameter), which significantly improved the model fit (ΔfREML = –5695.88, Δdf = 0, AIC difference = –11400.42, $p < .001$), highlighting the importance of accounting for temporal dependencies in eye-movement data. See Appendix E.3 for the model summary.

First of all, as expected, all groups successfully integrated the linguistic input, evidenced by increased looks to the target object in all conditions well above chance (Figure 3.9). This confirms general task engagement and comprehension across groups. The combined analysis, however, revealed an interesting pattern of group differences in the predictive versus non-predictive conditions. Both L1 Italian typical readers and L2 Italian/L1 Tyrolean speakers demonstrated significantly greater target-directed looks in the predictive condition compared to the non-predictive one. GAMM-derived difference curve analyses revealed significant effects from 306 ms to 709 ms post-target onset (effect size = 0.603) for the Italian L1 group and from 426 ms to 723 ms (effect size = 0.552) for the L2 group (see Figures 3.10a and 3.10c, respectively). These time-locked differences suggest that both groups were able to exploit predictive cues during real-time language processing,

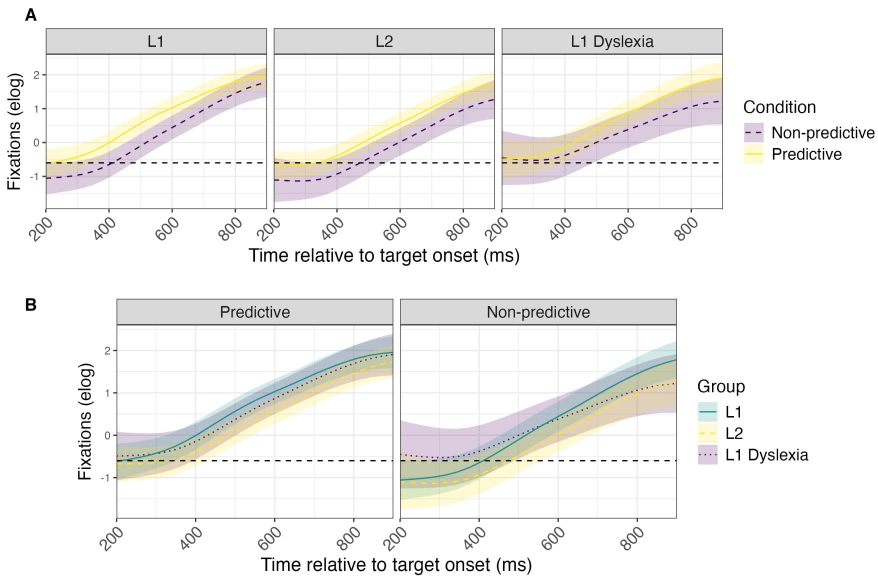
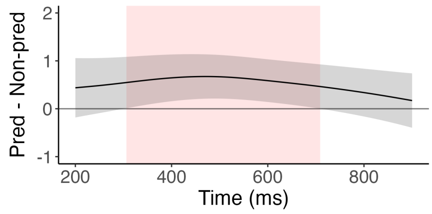


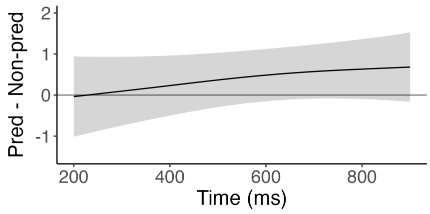
Figure 3.9: GAMM model estimates of looks to the target over time. (A) Comparison by condition for each group. (B) Comparison by group for each condition. The x-axis represents time relative to the target word onset (in ms), and the y-axis shows the proportion of looks directed to the target word transformed into empirical logit.

observed after the acoustic onset of the target. Interestingly, L1 participants with dyslexia did not exhibit significant differences between conditions over time (Figure 3.10b), indicating the absence of grammatical gender-based facilitation effects during target resolution in this group.

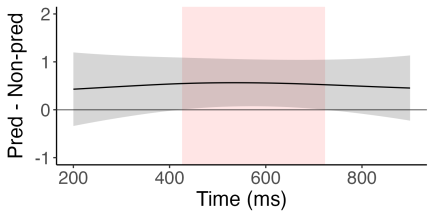
To assess whether the onset of preferential looks to the target in the predictive condition (relative to the non-predictive condition) occurred earlier in the Italian L1 group compared to the Tyrolean group, I conducted a divergence point analysis using 2000 bootstrap iterations. Divergence points were estimated separately for each group, and the group-level comparison was tested via bootstrapped permutation (see Section 2.2.5 for a detailed description of the method). This analysis did not detect a significant difference in divergence points between the two groups ($p = 0.7$; Figure 3.11). The estimated mean divergence point for the L1 typical readers was approximately 360 ms (95% CI = [300, 500] ms), while for the L2 group, it was 380 ms (95% CI = [300, 550] ms), resulting in a non-significant mean difference of approximately 29 ms (95% CI = [-200, 200] ms). These results



(a) L1 Italian typical readers



(b) Participants with dyslexia



(c) L2 Italian L1 Tyrolean speakers

Figure 3.10: Panels show differences in empirical logit-transformed fixation proportions over time between the predictive and non-predictive conditions for (A) L1 Italian typical readers, (B) participants with dyslexia, and (C) L2 Italian L1 Tyrolean speakers. The x-axis represents time relative to target word onset (0 ms). Red-shaded areas indicate time intervals where the difference between conditions is statistically significant. Grey bands represent standard errors around the estimated difference smooths.

suggest that the timing of preferential looks to targets in the predictive condition compared to the non-predictive one did not differ significantly between L1 and L2 participants.

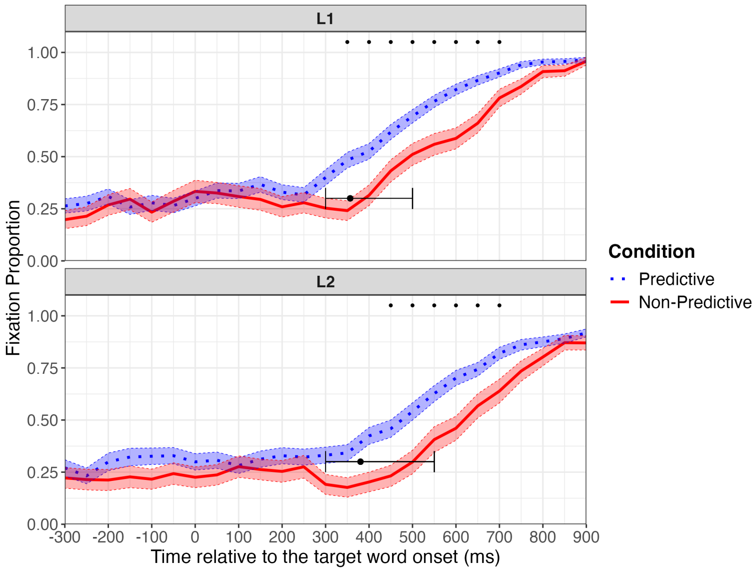
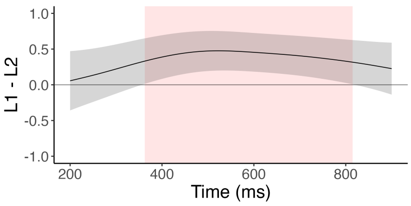


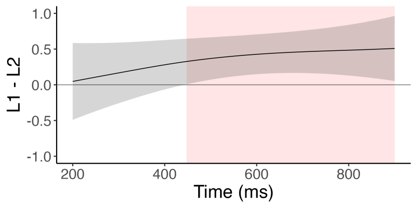
Figure 3.11: Divergence point analysis of mean fixation proportions over time for the predictive and non-predictive conditions. Fixation proportions to the critical word are plotted over time relative to the target word onset (0 ms), separately for L1 and L2 groups.

3.2.3.2 Combined Predictive vs. Non-Predictive Analysis: Between-Group Comparisons

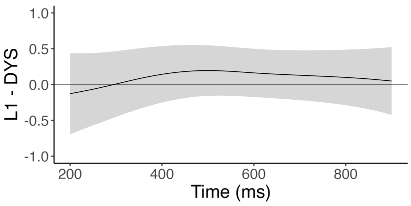
When comparing target fixations across groups within each condition, several significant differences emerged. In the predictive condition, the L1 Italian group demonstrated significantly higher target fixation proportions than the L2 group from approximately 363 ms to 815 ms (effect size = 0.417) (Figure 3.12a), but not compared to the dyslexia group (Figure 3.12c). The non-predictive condition also showed significantly more looks to the target in the L1 typical readers group compared to the L2 group from 447 ms until the end of the time window (effect size = 0.441) (Figure 3.12b), and no difference between L1 typical readers and readers with dyslexia (Figure 3.12d). No difference was observed betweent the L2 and dyslexia groups in neither condition, as indicated by confidence intervals that consistently included zero throughout the analysis window (Figure 3.12e-f).



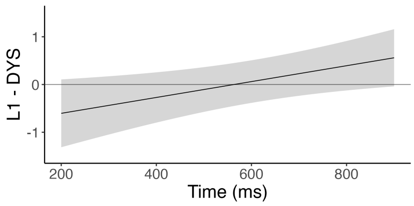
(a) L1 Italian vs L1 Tyrolean - Predictive condition



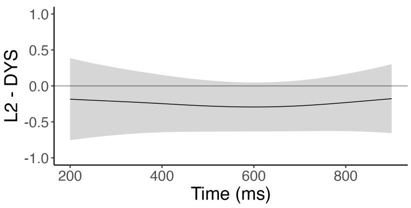
(b) L1 Italian vs L1 Tyrolean - Non-predictive condition



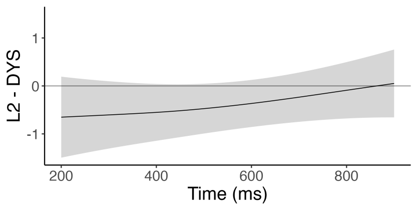
(c) L1 Italian vs Dyslexia - Predictive condition



(d) L1 Italian vs Dyslexia - Non-predictive condition



(e) L1 Tyrolean vs Dyslexia - Predictive condition



(f) L1 Tyrolean vs Dyslexia - Non-predictive condition

Figure 3.12: Between-group comparisons of target fixations across experimental conditions. Each panel shows the estimated difference in empirical logit-transformed fixation proportions over time, with 95% confidence intervals. Positive values indicate greater target fixations by the first group mentioned in each comparison. Pink-shaded regions indicate time intervals with statistically significant group differences. (a-b) L1 Italian typical readers vs L1 Tyrolean speakers. (c-d) L1 Italian typical readers vs participants with dyslexia. (e-f) L1 Tyrolean speakers vs participants with dyslexia.

At first glance, this might seem to contradict the earlier within-group time-course findings, where both L1 and L2 groups showed enhanced fixations to the target in the predictive condition compared to the non-predictive one, indicative of successful use of predictive cues. The current between-group contrasts, however, reflect absolute differences in target fixations. In other words, while both L1 and L2 speakers seem to exploit grammatical cues during the target recognition phase, the L2 group consistently shows overall lower proportions of target fixations across all the conditions. This pattern is consistent with previous findings in Experiment 1a, which showed that during word recognition, L2 speakers, though they integrated well the upcoming linguistic input, demonstrated overall lower proportions of looks to the critical word in the target condition (see Section 2.2.6).

3.3 Discussion

Absence of anticipatory effects. The complete absence of anticipatory effects across all participant groups challenges the notion that grammatical gender cues trigger automatic and rapid prediction during spoken language comprehension. This finding stands in contrast to some previous studies reporting early anticipatory effects based on gender-marked determiners in L1 speakers, as demonstrated by ERP effects on articles and adjectives (e.g., Ito et al., 2020; Otten & Van Berkum, 2009; Otten et al., 2007; Wicha et al., 2004) and anticipatory eye movements during spoken comprehension (Bosch & Foppolo, 2022; Huettig & Brouwer, 2015), although some studies report null results (Kochari & Flecken, 2019). Evidence in L2 speakers is more mixed: some studies report no anticipatory use of gender cues (Lew-Williams & Fernald, 2010), whereas others find predictive effects that are conditional on factors such as proficiency, lexical gender knowledge, or cross-linguistic similarity (Bosch & Foppolo, 2022; Dussias et al., 2013; Hopp, 2013, 2018). Importantly, research on individuals with dyslexia remains extremely limited, with only one study to date (Huettig & Brouwer, 2015) demonstrating that predictive use of gender cues is possible but substantially delayed. In the present study, no group exhibited significantly increased fixations to a target referent prior to its acoustic onset in the predictive condition.

Post-onset facilitation during word recognition. Although prediction was not observed, robust facilitation emerged in the resolution window (approximately 300–700 ms post-onset) for both L1 Italian and L1 Tyrolean/L2 Italian groups. In this time window, participants showed significantly greater looks to the target in the predictive compared to the non-predictive contexts, indicating that grammatical gender information supported more efficient word recognition.

On one view, these effects might reflect *ease of integration*: predictable words are easier to combine with prior context, without requiring pre-activation (Kutas & Federmeier, 2011; Luke & Christianson, 2016; Pickering & Gambi, 2018; Van Petten & Luka, 2012). However, as the sentences in the present experiment carried no semantic context to integrate beyond the grammatical gender of the determiner, a pure integration account is insufficient. An alternative account is that these effects reflect *prediction-facilitated lexical access*, where gender-marked determiners pre-activated features of the upcoming noun, but given the short time window, such anticipatory processes may have occurred too rapidly to be detected before target onset, leading to faster recognition once the noun was heard.

The current findings extend these effects to highly proficient L2 speakers, who also took advantage of morphosyntactic cues in real time, contrary to earlier reports of null effects in L2 learners (Lew-Williams & Fernald, 2010). This difference may lie in participant profiles. While in Lew-Williams and Fernald (2010) study L2 participants had on average five years of L2 instruction, L2 speakers in my study had been exposed to Italian for a mean of 19 years (range: 13–30). This interpretation was further supported by the divergence point analysis, which found no significant difference in the timing of divergence between the L1 and L2 groups ($p = .70$). Nevertheless, similar to Experiment 1a findings, the L2 group exhibited a shallower target preference, distributing more fixations across unrelated items and background despite orienting to the target at a similar time as L1 speakers. This pattern has been reported in other eye-tracking studies, where L2 learners' gaze behavior is characterized by reduced certainty and more diffuse fixations relative to natives (e.g., Corps et al., 2023; Mitsugi, 2020). These findings suggest that lower fixation proportions do not indicate failure to identify the target, but rather reflect the additional cognitive demands of L2 processing, which can lead to less sustained attention on the target once it has been selected.

Crucially, the facilitation effects observed in the neurotypical L1 and L2 groups were entirely absent in participants with dyslexia, who showed no difference between predictive and non-predictive conditions in the resolution window. Thus, while the temporal window might have been too short to reveal clear anticipatory looks, the observed post-target facilitation suggests that gender-marked articles prepared the processing system for more efficient lexical access in both L1 and L2 processing, but not in dyslexia, where morphosyntactic integration appears to be impaired.

Dyslexia and the absence of facilitation. The complete absence of facilitation effects in participants with dyslexia indicates they struggle to exploit grammatical gender cues during spoken language processing. The finding aligns with extensive evidence that individuals with dyslexia experience difficulties with morphological and morphosyntactic processing that extend well beyond written language. As reviewed in Section 1.4.2, adults with dyslexia show persistent impairments in morphological awareness tasks involving both inflectional and derivational processes with novel words (Melloni & Vender, 2022; Vender & Delfitto, 2024), as well as difficulties comprehending complex syntactic structures such as passive constructions, relative clauses, and clitic pronouns (e.g., Cardinaletti & Volpato, 2015; Marotta, 2022; Stella & Engelhardt, 2019; Wiseheart et al., 2009). These difficulties are particularly pronounced for structures involving long-distance dependencies or requiring the maintenance of grammatical features across intervening material (Arosio et al., 2016; Guasti et al., 2015). The present results extend this pattern to the online processing of grammatical gender agreement during spoken comprehension, demonstrating that morphosyntactic difficulties in dyslexia are not confined to offline metalinguistic tasks or written language, but also manifest during implicit, real-time integration of gender cues in the spoken modality.

Unlike Huettig and Brouwer (2015), who found delayed but detectable anticipatory effects in Dutch-speaking adults with dyslexia, our participants showed no such effect, even post-onset. Several factors may contribute to this difference. One possible explanation lies in the short determiner–noun interval of our paradigm, which may have imposed timing constraints that prevented dyslexic participants from generating or integrating predictive cues. In fact, Huettig and Brouwer (2015) proposed that predictive differences between groups are magnified under more demanding conditions. However, this timing account alone is insufficient, given that the neurotypical groups successfully integrated gender information in the same time window. A more likely explanation is that the integration of grammatical gender cues during real-time comprehension places particularly high demands on the morphosyntactic processing system, which is known to be impaired in dyslexia. Moreover, dyslexia has been linked to reduced verbal working-memory capacity, which can limit the availability of resources for prediction (Huettig & Janse, 2016; Otten & Van Berkum, 2009). It should be acknowledged, however, that working memory accounts are most compelling over longer processing windows, and the short determiner–noun interval of the present paradigm makes it less likely that working memory load per se was the primary bottleneck. Rather than pre-activating candidates, dyslexic individuals may also adopt a "wait-and-see" strategy, focusing on integration at noun onset. This aligns with the shallow processing pattern hypothesised by Engelhardt et al. (2021), where comprehenders showed no evidence of predictions in less constraining contexts. It should be noted that in spoken Italian, morphosyntactic information is conveyed through phonological form — the gender contrast between *il* and *la* is realized as a phonological distinction. It is therefore difficult to fully disentangle whether the reduced use of gender-marked determiners in the dyslexia group reflects a deficit at the level of morphosyntactic representations per se, or whether limitations in encoding and maintaining the phonological signal that realizes agreement may additionally constrain access to that morphosyntactic information during real-time processing. Rather than treating these as fully separable sources of difficulty, it may be more accurate to consider them as interacting: phonological processing limitations may constrain the speed or reliability with which morphosyntactic cues become available for integration during online comprehension.

Taken together, the present findings provide evidence that morphosyntactic processing difficulties in dyslexia extend to the rapid integration of grammatical gender cues during spoken language comprehension. Whether this deficit stems primarily from impaired morphosyntactic representations, limitations in phonological encoding of agreement cues, slower lexical access, or strategic adaptations such as "wait-and-see" processing (with working memory playing a more indirect role) remains an open question requiring further investigation. What is clear, however, is that individuals with dyslexia do not exploit gender-marked articles to facilitate lexical access in the same manner as neurotypical speakers.

Cognate status and cross-linguistic influence. The study also investigated whether cross-linguistic gender congruency between Italian and Tyrolean modulated predictive processing in bilinguals. The presence of gender-consistent or gender-conflicting determiners across the two languages was hypothesized to influence target identification. However, the absence of cross-linguistic gender congruency effects in the L2 group suggests that participants did not show differential processing based on cross-linguistic gender alignment in real-time comprehension. That is, target fixations were not modulated by whether the gender-marked articles in Italian and Tyrolean were congruent or incongruent. However, given that no anticipatory use of gender cues was detected even in Italian L1 speakers, the present null effect should be interpreted cautiously.

Several possible explanations may account for this null result. First, L2 speakers may not automatically activate L1 lexical representations, including grammatical gender features, during L2 processing. As reviewed in Section 1.4.1, many models of the bilingual lexicon assume broadly non-selective access across languages (e.g., Dijkstra & Van Heuven, 2002; Duyck, 2005; Marian et al., 2003; Schwartz et al., 2007), yet cross-linguistic information can be strategically inhibited when it is unhelpful for spoken comprehension (e.g., De Groot & Christoffels, 2006; B. J. Levy et al., 2007; Linck et al., 2009; Mercier et al., 2016; Misra et al., 2012). In our case, this may have reflected task-induced suppression of Tyrolean features in an Italian-only setting as the entire procedure, except for the word/non-word reading task, was conducted in Italian. Moreover, because our design included both congruent items (where cross-linguistic transfer could, in theory, facilitate processing) and incongruent items (where it would hinder processing), participants may have learned to avoid relying on L1 gender cues altogether. Future work could test whether effects would emerge in a paradigm with only congruent items, where reliance on L1 information could not lead to interference. Such a finding would support the view that bilingual lexical activation is not entirely automatic but is instead adaptable to task context and cue reliability. Second, cross-linguistic activation may emerge only at particular levels of proficiency or dominance. For instance, Ito et al. (2018) reported that highly proficient Japanese learners of English (mean exposure ≈13 years) predicted upcoming targets when comprehending in English but did not activate their phonological competitors in their L1, Japanese. Our L2 group, with a mean of 19 years of exposure to Italian, showed a parallel pattern: strong prediction of upcoming high CP Italian targets (Experiment 1a) but no evidence of cross-linguistic grammatical gender activation (Experiment 1b). At the same time, other studies point to proficiency effects in the opposite direction, linking higher L2 proficiency to greater L1 co-activation in some semantic tasks whereas lower proficiency – to L1 inhibition (e.g., Branzi et al., 2021). In the present dataset, the complete absence of cross-linguistic gender-congruency effects precluded an empirical test of proficiency interactions. Finally, given the asymmetry between the Italian (masculine/feminine) and Tyrolean (masculine/feminine/neuter) gender systems, participants may have implicitly learned that cross-linguistic gender predictions are unreliable and therefore avoided such strategies in real time. More generally, as mentioned above, comprehenders might be less likely to deploy predictive cues whose reliability is low across contexts. Together, these findings suggest that cross-linguistic grammatical gender alignment does not always take place in online comprehension, even when morphosyntactic features such as grammatical gender are available in both languages. The absence of cross-linguistic congruency effects in this study warrants further investigation using different language pairs, varying language proficiency levels and degrees of L1-L2 overlap, as well as experimental paradigms that more directly cue L1 activation.

In addition to grammatical gender congruency, the study examined whether lexical overlap across languages, in the form of cognates, modulated spoken language processing in bilingual participants. Contrary to expectations, no reliable cognate effects were observed in the L2 group across any condition, i.e., the presence of cognate words did not modulate fixation behaviour in any way. These results suggest that, under the present experimental conditions, cognate status did not influence predictive or integrative mechanisms in L2 speakers.

Theoretical and practical implications. The findings presented here offer several important insights. First, they caution against assuming that morphosyntactic anticipatory processing is automatic across populations. While most prior studies in L1 comprehension report anticipatory use of grammatical gender cues (e.g., Hopp, 2013; Lew-Williams & Fernald, 2010; Otten & Van Berkum, 2009; Otten et al., 2007; Wicha et al., 2004), the present data did not reveal such effects. This discrepancy may partly reflect methodological factors, in particular the short determiner–noun interval of the paradigm, which could have limited the time available for anticipatory processing. It is worth noting that prior studies reporting anticipatory effects typically employed paradigms that extended the determiner–noun interval through intervening adjectives or words when the language permits, or through prosodic manipulations such as guessing-game intonation contours (Bosch & Foppolo, 2022; Huettig & Brouwer, 2015), thereby providing participants with substantially more time to generate predictions. The present paradigm used Italian sentences, which offer far fewer options for lengthening the interval without simultaneously introducing additional gender cues (e.g., grammatical gender marking on intervening adjectives). The absence of anticipatory effects may therefore partly reflect the limits of prediction under these more constrained conditions, rather than solely a limitation of the current design. Importantly, however, facilitation effects were observed in the resolution window for the neurotypical groups, indicating that gender cues were successfully integrated once the target word became available. The complete absence of such facilitation in the dyslexia group, however, cannot be attributed to timing constraints alone, as the neurotypical groups successfully integrated gender information in the same time window.

Second, the results confirm that, in the present task, highly proficient L2 speakers can integrate morphosyntactic information, though with somewhat reduced efficiency relative to native speakers. This pattern suggests that L2 processing in this domain may not be fundamentally different, but rather subject to quantitative limitations (Kaan, 2014).

Third, the absence of facilitation effects in participants with dyslexia is in line with prior evidence of morphosyntactic processing difficulties in this population. As discussed in Section 1.4.2, research has documented deficits in morphological awareness tasks involving inflectional and derivational processes with novel words in both children and adults with dyslexia (Melloni & Vender, 2022; Vender & Delfitto, 2024), as well as difficulties with complex syntactic structures such as relative clauses, passive constructions, and clitic pronouns (Cardinaletti & Volpato, 2015; Marotta, 2022; Stella & Engelhardt, 2019). The present findings demonstrate that these difficulties also manifest during implicit, real-time integration of grammatical gender cues in spoken comprehension. From a practical perspective, these findings support the development of targeted pedagogical interventions that train not only phonological and orthographic skills but also the use of morphosyntactic cues during comprehension.

Finally, the absence of cross-linguistic congruency and cognate effects in the current study suggests that bilinguals may engage primarily in language-selective morphosyntactic processing under the present task conditions. One possibility is that the relatively high proficiency of participants, combined with an Italian-only task context, reduced the likelihood of cross-language activation. This finding indicates that cross-linguistic interaction is not automatic but modulated by factors, such as task demands, proficiency levels, or the structural similarity between languages. Future studies could test this hypothesis by examining participants with varying degrees of language dominance and proficiency. Additionally, extending this paradigm to language pairs with greater morphosyntactic overlap or typological similarity may clarify whether structural similarity facilitates cross-linguistic activation. Finally, incorporating neurophysiological measures (e.g., EEG, MEG) could provide more sensitive measures of subtle cross-linguistic effects not captured by eye-tracking alone.

Chapter 4

Experiment 2: Predictive Processing in English: Contextual, Semantic, and Phonological Effects

The findings from Experiment 1a conducted in Italian (see Chapter 2) provided compelling evidence for contextual prediction, with all participant groups demonstrating anticipatory fixations to highly predictable target words, while revealing some differences in the temporal dynamics of predictive processing between L1 and L2 participants. Notably, no evidence emerged for phonological form pre-activation, suggesting that phonological prediction may be limited or require very specific conditions (Ito & Husband, 2017). However, several important questions remained unresolved, and the opportunity arose to address these limitations through a second experiment conducted in English at the University of East Anglia, UK.

One of the primary motivations for the current experiment was to increase the dyslexia sample size, addressing the small number of participants in the dyslexia group in Experiment 1, which contributed to increased statistical variability. Another aim was to enhance the generalizability of the findings by validating them in a different linguistic context. To this end, Experiment 2 adopted the visual world paradigm and stimulus materials developed by Ito and Husband (2017), used with the authors' permission. While the present study was not designed as a direct replication of Experiment 1a, it addressed the same core research questions while incorporating several methodological refinements. These additions included: (1) the investigation of semantic competitor activation, which was not addressed in Experiment 1 and allowed for the analysis of pre-activation patterns of semantically related words, thereby providing a window into whether contextual prediction extends beyond the specific target to activate related concepts; (2) the inclusion of a separate condition to control for lexical activation in unpredictable contexts, ensuring that any observed competitor effects reflected genuine prediction rather than general differences in visual attention or lexical processing; (3) a comprehensive individual differences test battery to better characterize participant populations through standardized measures.

The original study by Ito and Husband (2017) represents an investigation into the strength and temporal characteristics of contextual, semantic, and phonological prediction, using a visual world paradigm with highly predictable sentence contexts. Their key finding that participants showed robust semantic competitor activation but no evidence of phonological prediction during spoken language comprehension challenged earlier ERP findings suggesting phonological form pre-activation (DeLong et al., 2005; Martin et al., 2013) and contributed to the debates about the conditions under which phonological prediction occurs. These findings, however, remain controversial (see Section 1.3.3 for a review of phonological prediction studies). Importantly, the original Ito and Husband (2017) study tested only native English neurotypical speakers, leaving open questions about how these prediction patterns might manifest across different populations. By replicating their paradigm while systematically examining individual differences across L1, L2, and dyslexic participant groups, Experiment 2 aimed to extend these findings beyond the original neurotypical L1 population and investigate whether population differences in contextual and phonological prediction align with patterns observed in Experiment 1. Additionally, the inclusion of a comprehensive individual differences battery would provide insights into the mechanisms underlying any observed group differences in prediction patterns.

To ensure both methodological rigor and participant comfort, two changes were implemented to the original Ito and Husband's study. First, the preview time was shortened from 3000 ms before sentence onset to 2000 ms. This adjustment helped reduce the overall eye-tracking task duration to approximately 30 minutes, minimizing participant fatigue and attention decrements that could compromise data quality. Importantly, a 1000–2000 ms preview window is generally regarded as sufficient for visual encoding while preserving the validity of prediction measurements (Li et al., 2022). Generally, very short previews ($<$ 500 ms) risk insufficient encoding of the visual scene, obliterating any possible phonological effects (Huettig & McQueen, 2007), while longer previews mainly increase experiment time but don't necessarily yield stronger prediction effects.

Second, the speech rate was increased from the original 2.5 syllables/second used in Ito and Husband to 3.5 syllables/second. Speech rate has been shown to systematically modulate anticipatory eye movements in the visual world paradigm, with prediction effects following a non-linear relationship with rate (Fernandez, Hadley, et al., 2025; Fernandez et al., 2020; Huettig & Guerra, 2019). In particular, intermediate rates (approximately 3.5–4.5 syllables/second) have been associated with more robust anticipatory effects, compared to slower or faster speech. Thus, the selected rate of 3.5 syllables/second provided tighter time-locking between attentional shifts and linguistic processing compared to slower rates, while avoiding speech rates that may place excessive perceptual demands, particularly on non-native listeners. Although English is traditionally classified as a stress-timed language (e.g., Setter & Sebina, 2017), speech rate in visual world studies is most commonly quantified in syllables per second, which allows for comparability across studies and across participant populations. Moreover, to ensure that the speech rate reflected the acoustic input actually perceived by listeners, rate estimation was based on broad phonetic transcriptions of the recorded utterances rather than orthographic or dictionary forms, similarly to Experiment 1a (see Section 2.1.2.2).

Experiment 2 addressed four primary research questions:

1. Are phonological features of an upcoming highly predictable word retrieved before its acoustic onset during spoken sentence comprehension in English?
2. Does contextual prediction extend to semantically related words during spoken language comprehension?
3. Do adults with developmental dyslexia differ from typical readers in their anticipatory spoken language processing in English across different linguistic levels? If so, are prediction differences in speakers with dyslexia attributable to general processing speed, attention symptoms, reading history difficulties, or vocabulary knowledge and print exposure?
4. How does second language status affect the time course of predictive processing during English comprehension across different linguistic levels?

4.1 Method

4.1.1 Participants

Ninety-five participants were recruited from the University of East Anglia campus and completed the study. Recruitment was conducted through multiple channels, including university mailing lists, campus advertisements, and the psychology department participant pool. The study included three participant groups: (1) L1 English speakers with typical reading development (the control group, n = 39), (2) L1 English speakers who had an official dyslexia diagnosis by a qualified professional and were registered with the University of East Anglia's disability services for learning disabilities (n =

29), and (3) L2 English speakers whose native language was one of the Romance languages and who had acquired English as a second language outside the family environment (n = 25). The L2 group comprised speakers of Italian (n = 5), Spanish (n = 8), French (n = 1), Portuguese (n = 3), and Romanian (n = 8). In the overall design of the dissertation, Experiment 1 compares L1 Italian speakers (a Romance language) with L1 Tyrolean speakers (a Germanic language), whereas Experiment 2 reverses this typological configuration by testing L1 English speakers (Germanic) together with L2 English speakers whose L1 is Romance. For the L2 group, detailed information about language background, age of acquisition, patterns of daily language use, and academic and professional exposure to English was collected using the Language Experience and Proficiency Questionnaire (LEAP-Q; see Sections 4.1.2.1 and 4.2.1). All L2 participants were resident in the UK at the time of testing and were enrolled as university students.

Table 4.1: *Demographic characteristics of study 2 participants by group*

	English L1		L2
Characteristic	Typical	Dyslexia	
Sample size (n)	39	29	25
Gender (%male)	25.6	17.2	29.2
Age (years)	19.87 (1.78)	23.79 (7.06)	22.96 (4.73)

Note. Values show Mean (Standard Deviation).

Following data cleaning procedures (detailed in Section 4.2.2), two participants were excluded from the final analyses. One participant from the dyslexia group was excluded due to insufficient valid eye-tracking data following preprocessing. One participant from the L1 typical reader group was excluded due to extreme outlier performance on multiple measures from the individual differences battery. Thus, the final dataset consisted of a total of 93 participants. Table 4.1 presents demographic characteristics for each group, including age and gender distribution.

Inclusion and exclusion criteria. All participants met the inclusion criteria of normal or corrected-to-normal vision and hearing. For the dyslexia group, participants were additionally required to provide official diagnostic documentation. For the L2 group, participants were required to have a Romance language as their L1, spoken in the family from birth, and to have learned English after the age of 3 years and outside the family environment to ensure they were not simultaneous bilinguals.

4.1.2 Materials

4.1.2.1 Preliminary Measures

A comprehensive battery of standardized assessments was administered to all participants prior to the eye-tracking experiment. This battery was designed to characterize participant groups across multiple cognitive domains, to ensure correct group classification, and to investigate potential relationships between individual differences and predictive processing abilities. All standardized assessments were administered and scored in accordance with the procedures specified in their respective test manuals, unless otherwise noted.

Rapid Automatized Naming (RAN). All participants completed letter and number RAN tasks using the Comprehensive Test of Phonological Processing (CTOPP-2; Wagner et al., 2013). Both tasks required participants to name four rows of nine stimuli (letters or numbers) sequentially aloud as quickly and accurately as possible. Participants completed one practice row before each task. The score was the total time required to complete each array in seconds divided by the number of items minus the number of errors (misnaming, omissions), with higher scores indicating slower processing speed. Letters and numbers were presented in 20-point Arial font on off-white A4 paper. The reliability of the CTOPP-2 subtests, including RAN tests, has been reported by average internal consistency that exceeds .80. RAN tasks are widely recognized as sensitive measures of processing speed deficits in dyslexia and were expected to differentiate the dyslexia group from controls (Araújo et al., 2011; Denckla

& Rudel, 1976; Georgiou et al., 2018; Norton & Wolf, 2012; Wolf & Bowers, 1999).

Adult Reading History Questionnaire (ARHQ). Participants completed a self-report screening tool designed to assess dyslexia risk in adults (Lefly & Pennington, 2000). The ARHQ evaluates reading history, current reading habits, and reading-related difficulties across the lifespan (for example, "how much difficulty did you have learning to spell in elementary school" or "how much reading do you do for pleasure"). While not constituting a formal diagnosis, the ARHQ provides a standardized measure of reading history and reading habits. Participants responded to each question using a Likert scale from 0 to 4, and the total score was determined by summing all the points. Higher scores indicate greater reading difficulties and dyslexia risk. The questionnaire was used as one of the measures to confirm group classification and to identify any potential undiagnosed reading difficulties in the control group.

Peabody Picture Vocabulary Test-4 (PPVT-4). Receptive vocabulary was assessed using the PPVT-4 (Dunn & Dunn, 2007), a widely used standardized measure of vocabulary knowledge. In this task, the experimenter named target words orally, and participants selected from four pictures the one that best illustrated the word's meaning. The test uses an adaptive procedure, beginning with easier items and continuing until a ceiling is reached. Form A was administered to all participants. The reliability range for Form A is good to excellent, reported to be ranging from .89 to .97 across age groups (Dunn & Dunn, 2007). Beyond serving as a measure of receptive vocabulary, performance on the PPVT is also influenced by cumulative language and print exposure. Reading provides access to low-frequency and abstract vocabulary that is less common in spoken discourse, making receptive vocabulary size a useful proxy for reading experience (Cunningham & Stanovich, 1991; Stanovich, 1986).

Author Recognition Test (ART). Print exposure was measured using an updated version of the Author Recognition Test (Acheson et al., 2008) originally developed by Stanovich and West (1989). Participants are presented with a list of names on a single answer sheet and are asked to identify which of the listed names belong to real authors. The list contains 130 names in total: 65 genuine authors' names (e.g., Ernest Hemingway) and 65 foils (non-authors, e.g., Geoffrey Pritchett). Participants were instructed to tick only those names they were confident about and to avoid guessing, as selecting a foil resulted in a penalty. The ART is a standardized, indirect measure of print exposure, as individuals who read more are more likely to recognize authors' names. Scores reflect the number of correctly identified authors minus the number of incorrectly selected foils, providing a measure that controls for guessing.

Stroop color-word task. Executive control and inhibitory abilities were assessed using a computerized version of the Stroop task (MacLeod, 1991; Stroop, 1935). Participants viewed color words ("red," "blue," "green") presented in either congruent ink colors (e.g., "red" in red ink) or incongruent ink colors (e.g., "red" in blue ink) and were instructed to respond to the ink color while ignoring the word meaning by pressing designated keys ('b' for blue, 'r' for red, and 'g' for green). Each participant saw 36 trials, of which 12 were congruent and 24 incongruent. The dependent measure was the difference in reaction time between incongruent and congruent trials (Stroop interference effect), with larger differences indicating poorer inhibitory control. This measure was included given some theoretical proposals that prediction in language comprehension may involve inhibitory mechanisms to suppress competing lexical alternatives, especially when predictions are violated (Kim et al., 2023; Kukona et al., 2016), and evidence suggesting that individuals with dyslexia show difficulties in executive and inhibitory control (Barbosa et al., 2019; Reiter et al., 2005).

Language Experience and Proficiency Questionnaire (LEAP-Q). All participants from the L2 group completed the LEAP-Q (Marian et al., 2007), a comprehensive assessment of language background and self-rated proficiency. Similar to Experiment 1, the questionnaire collected information about language acquisition history, current language use patterns, and self-assessed proficiency across four modalities (speaking, listening, reading, writing) for all known languages. This information was important for characterizing L2 participants' English proficiency and language background.

Attention Deficit Hyperactivity Disorder (ADHD) symptom assessment. Given the documented comorbidity between dyslexia and ADHD (with a ∼30% co-occurrence probability) (Pennington, 2006), participants completed a brief self-report screening of ADHD symptoms using the Adult ADHD Self-Report Scale (ASRS-v1.1; Kessler et al., 2005). This instrument assesses current inattention and hyperactivity symptoms (e.g., "How often do you feel restless or fidgety?"). The measure was included to control for potential confounding effects of attentional difficulties on predictive processing performance.

Following data collection, the relationships between individual difference measures and predictive processing outcomes (divergence points in the eye-tracking data) were examined through correlation analyses.

4.1.2.2 Task Overview and Experimental Conditions

The experimental materials were taken from the original study by Ito and Husband (2017) with permission of the authors, employing a visual world paradigm that systematically manipulated both sentence predictability and object relationships to the target word. The experimental task involved participants viewing arrays of four pictures while listening to English sentences containing highly predictable words. Their task was to click on a picture of the object mentioned in the sentence or on the background if none of the pictures were mentioned as fast as possible. The study implemented a 2×4 factorial design with the following factors:

Factor 1: Sentence Predictability (within-subjects)

- **Predictable sentences**: Highly constraining contexts with 95% mean cloze probability (e.g., "Instead of sending an email, Ellie decided to write a *letter* to her sister," see Figure 4.1 for the example of the visual display.)
- **Unpredictable sentences**: Low-constraint contexts using the same target words (e.g., "If there is one, click on the picture of the *letter*.")

Factor 2: Object Type (within-subjects)

- **Target**: the object explicitly mentioned in the sentence and highly predictable in constraining contexts (e.g., *letter*)
- **Semantic competitor**: an object semantically related to the predicted target, used to test whether prediction spreads beyond the specific lexical item (e.g., *postbox*)
- **Phonological competitor**: an object sharing onset overlap with the target, used to assess phonological pre-activation (e.g., *lettuce*)
- **Unrelated**: an object unrelated both semantically and phonologically, serving as a baseline control (e.g., *crab*)

Additionally, Participant Group served as a between-subjects factor with three levels: L1 English speakers, L1 English speakers with dyslexia, and L2 English speakers.

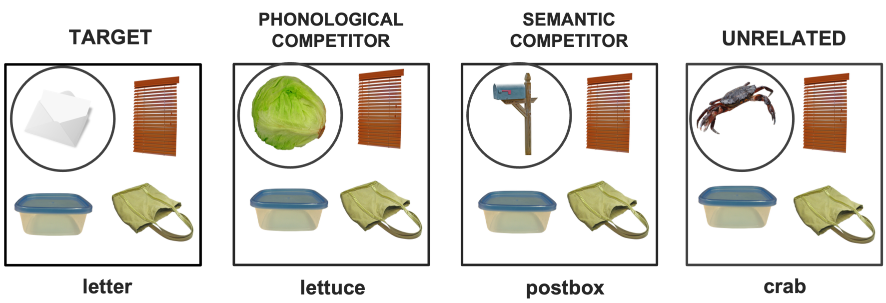


Figure 4.1: Example visual displays for the predictable sentence "Instead of sending an email, Ellie decided to write a *letter* to her sister," and for the unpredictable sentence "If there is one, click on the picture of the *letter*."

The full corpus initially comprised 56 predictable experimental sentences and 56 unpredictable experimental sentences. During pre-analysis checks, two pairs were identified as unsuitable and were removed, resulting in a final set of 54 experimental pairs (112 experimental sentences overall), plus two practice trials (see Appendix C for the full list). One item was excluded due to an unintended form-based relationship between the verb *unlock* in the sentence "Alice tried to unlock the door but realized she had forgotten her keys and couldn't get in" and the image used as the semantic competitor (a padlock). This unintended phonological overlap likely biased anticipatory eye movements, leading participants to fixate the competitor more than the actual target (keys). A second item was discarded because one of the distractor images (ginger) elicited abnormally high fixation proportions, most likely due to its visual salience. All remaining predictable sentences contained highly predictable target words, with cloze probability normed at 95% (range: 80-100%) based on the original Ito and Husband (2017) validation. Sentences in the predictable block were variable among themselves in terms of syntactic structure, tense, length, and the place of the target. A key design difference from Experiment 1a concerned the placement of critical targets. In Experiment 1a, sentence structures were uniform and targets always occurred in sentence-final position. This consistent placement created predictable temporal windows for anticipatory processing, which was advantageous for maximizing experimental control and comparability across items. In the current Experiment 2, by contrast, sentence structures followed the original design of Ito and Husband (2017), in which the target words appeared in different positions within the sentence. This allowed for examining predictions in scenarios where the target's location is not fixed. The sentences consisted of a mean of 20.6 syllables (SD = 4.4, range = 12 - 39 syllables), and their mean duration was 5.91 s (SD = 1.2). The unpredictable sentences maintained the same target words but embedded them in contexts that provided no semantic constraint: "If there is one, click on the picture of the [target]."

Each trial presented four pictures arranged in a 2×2 grid on the computer screen. The critical object (target, semantic competitor, phonological competitor, or unrelated, depending on condition) appeared in one quadrant, while three fixed distractor objects occupied the remaining positions (see Figure 4.1). To ensure that competitor effects reflected genuine prediction rather than general plausibility preferences, all objects except the target were implausible within the predictable sentence context. This ensured that any fixation bias toward semantic or phonological competitors would be driven specifically by their semantic or phonological overlap with the predicted target word. Unlike Experiment 1a, Experiment 2 employed coloured images, following the original materials provided by Ito and Husband (2017). No formal screening for colour vision deficiencies was administered. However, objects were identifiable based on their shape and canonical visual properties, and experimental conditions were not defined by colour contrasts. Even if some items may benefit from canonical colouring, colour cues were not systematically aligned with any object type or condition, making it unlikely that colour vision differences could account for the observed competitor effects.

Semantic competitors had varied relationships with target words, including belonging to the same semantic category (e.g., *nose-ear*), part-whole relationships (e.g., *car-tire*), and frequent co-occurrence patterns (e.g., *tea-lemon*). Phonological competitors shared an average overlap proportion of 0.49 (SD = 0.16) with target words at onset. The overlap was computed following Simmons and Magnuson (2018) as the number of identical phonemes divided by the average phoneme length of the two words. Importantly, overlap estimates were grounded in the phonetic realization of the critical words in the recorded auditory materials, rather than in abstract dictionary forms. The procedure followed the same approach adopted in Experiment 1a (see Section 2.1.2.2). See Appendix C.2 for full transcriptions of the sentence stimuli.

In addition to the experimental sentences, the design included 56 filler sentences, divided evenly between the two blocks (28 per block). Filler sentences in the predictable block had unpredictable contexts that did not constrain expectations toward any particular object (e.g., "Alex found a toad on his way home" presented with pictures of *a toad*, *a birdhouse*, *a boxing glove*, and *a screwdriver*). Filler sentences in the unpredictable block used the same neutral structure as the experimental unpredictable items ("If there is one, click on the picture of the..."). The proportion of trials displaying a mentioned object was carefully controlled. Within each set of 28 fillers, 21 sentences mentioned one of the four displayed objects, while 7 mentioned none of the pictured objects, yielding a 75%/25% structure. This structure was designed to balance the experimental logic: each experimental item had four possible versions (target, semantic competitor, phonological competitor, unrelated), but only the target version displayed the named object on screen — meaning only 25% of experimental trials showed the named object. By having 75% of fillers show a named object, the overall design resulted in 50% of all trials displaying a mentioned object throughout the experiment.

Eight experimental lists were created by crossing experimental conditions with block order (predictable first vs. unpredictable first), ensuring that participants encountered only one version of each item, with the specific condition determined by their assigned list. Critical objects appeared equally frequently in each screen quadrant across participants. The auditory materials for Experiment 2 consisted of original recordings provided by Ito and Husband (2017), recorded by a female native speaker of British English.

4.1.3 Apparatus

Participants' eye movements were recorded using an EyeLink 1000 Desktop Mount eye-tracker sampling at 500 Hz (SR Research Ltd., 2005). This represented a substantial upgrade from the Gazepoint GP3 HD eye-tracker (60 Hz, Gazepoint, 2024) used in Experiment 1. The higher sampling rate of the EyeLink 1000 provides enhanced temporal precision for measuring saccade onsets and fixation durations, though both systems provide adequate resolution for studying fixation patterns in the visual world paradigm (Raney et al., 2014).

The experiment was programmed using SR Research Experiment Builder software (Version 2.6.11; SR Research Ltd., 2018). Visual stimuli were presented on a 24-inch ASUS LCD monitor with a resolution of 1920 × 1080 pixels, while eye-movement data were recorded on a separate Dell OptiPlex 7050 host computer running the EyeLink system. Auditory stimuli were presented through Creative A60 2.0 external speakers. The speakers have a frequency response of 50–20,000 Hz, a total RMS output power of 4 W (2 × 2 W), and a signal-to-noise ratio of 75 dB. They were positioned symmetrically on either side of the monitor.

4.1.4 Procedure

All participants provided informed consent and received compensation in the form of course credits or £7 based on their preference. The experimental session was conducted individually in a quiet laboratory setting and lasted approximately 60 minutes. The study was conducted in accordance with the principles of the Declaration of Helsinki and the British Psychological Society Code of Human Research Ethics. Ethical approval was obtained from the University of East Anglia Research Ethics Committee, the host institution where data collection took place. All participants provided informed consent. Personal data were handled in compliance with the General Data Protection Regulation (GDPR) and were anonymised at the time of collection through the use of unique participant codes.

Upon arrival, participants completed a demographic questionnaire where they indicated their age, gender, native language(s), vision correction needs, and any diagnosed language or learning disorders. Participants assigned to the L2 group additionally completed the Language Experience and Proficiency Questionnaire (LEAP-Q; Marian et al., 2007), which was used to document English exposure and self-rated proficiency. Following these questionnaires, all participants completed the comprehensive individual-differences battery described in Section 4.1.2.1.

Following completion of the preliminary measures, the eye-tracking phase began. Participants were seated comfortably approximately 60 cm from the eye-tracker with their head positioned on a chin rest to minimize movement. The experimental room was kept at consistent lighting levels to ensure stable pupil tracking. Prior to each experimental session, a standard 9-point calibration procedure was conducted, during which participants fixated on calibration targets presented across the screen until the system achieved acceptable accuracy. Drift correction was implemented at the beginning of each trial using a central fixation point, ensuring maintained accuracy throughout the session.

Participants received instructions in English, presented both visually on the screen and aurally, explaining that they would view four pictures while listening to sentences. They were told that sometimes (but not always) one of the pictures would be mentioned in the sentence, and their task was to click on that picture as quickly as possible when it was mentioned. If none of the pictures were mentioned, they were instructed to click on the background instead. To get familiarized with the procedure, two practice trials were provided.

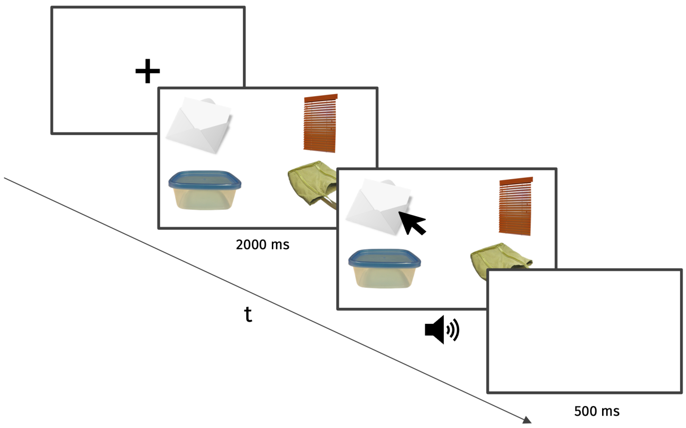


Figure 4.2: An example of the experimental procedure.

Each trial began with a central fixation cross appearing on screen for drift correction. If drift correction failed, recalibration was performed. Following successful drift correction, four pictures appeared on screen in the four quadrants. The audio sentence began playing 2000 ms after the picture onset. Participants clicked on the relevant picture when mentioned in predictable trials or clicked anywhere on the background in competitor and unrelated trials where the target was not present. Trials ended upon mouse click response (see Figure 4.2). Given the complex structure of the original experiment from Ito and Husband (2017), the clicking component was preserved to avoid potential programming errors that could arise from modifying the established protocol. Moreover, this design choice eliminated potential ambiguity about task expectations, as participants had an explicit action to perform when the target was mentioned.

The session concluded with a brief debriefing where participants could ask questions about the study and provide feedback about their experience. To reduce bias, the experimenter did not communicate any specific hypotheses during data collection.

4.1.5 Predictions

Based on the findings from Experiment 1a (Chapter 2), the theoretical framework outlined in previous chapters, and the results reported by Ito and Husband (2017), the following predictions were formulated:

Contextual and semantic prediction. All three groups were expected to show anticipatory fixations on highly predictable targets, consistent with extensive evidence that comprehenders routinely anticipate high-cloze targets in strongly constraining contexts (Altmann & Kamide, 1999; Borovsky et al., 2012; Ito et al., 2018) and Experiment 1a results. Beyond target prediction, strong semantic competitor effects were expected, replicating Ito and Husband (2017) while extending these findings to dyslexia and L2 populations. The available evidence, though limited, consistently indicates that prediction extends beyond single target items to include semantically related concepts (Federmeier & Kutas, 1999; Ito & Husband, 2017; Li & Qu, 2023; Li et al., 2022, see Section 1.3.1). I therefore predicted that all groups would show anticipatory fixations to semantic competitors relative to unrelated objects in predictable contexts. However, the timing and strength of these effects were expected to vary across populations. Compared to L1 typical readers, both dyslexia and L2 groups were predicted to show a delayed onset of semantic prediction. For dyslexia, this prediction was motivated by evidence that, despite relatively preserved performance in highly constraining contexts, predictive processing in dyslexia unfolds less incrementally and more slowly over time (Engelhardt et al., 2021). Adults with dyslexia show slower lexical retrieval, as indexed by reduced rapid automatized naming performance (Araújo et al., 2011; Carioti et al., 2021). While strong contextual constraints can support accurate target anticipation, the activation of semantically related alternatives depends on the efficiency with which the target word and its associated semantic network become active (Huettig et al., 2022). Because semantic competitor activation arises from spreading activation within this network, impairments in the speed of lexical access should manifest as delays in when this spreading reaches related concepts. Consequently, semantic competitor effects were expected to emerge later in this group, with potential convergence to control patterns as bottom-up input accumulates. For L2 speakers, slower and less automatic lexical access, weaker form–meaning mappings, and increased lexical competition typically delay the pre-activation of semantic information (Dijkgraaf et al., 2017; Martin et al., 2013; Weber & Cutler, 2004). Semantic competitor effects in the L2 group were therefore expected to be temporally delayed relative to L1 typical readers, while remaining qualitatively similar in direction.

Phonological prediction. Previous work has provided little evidence for phonological form pre-activation during spoken language comprehension. Both Ito and Husband (2017) and Experiment 1a failed to detect phonological competitor effects. At the same time, Ito's (2024) meta-analysis indicates that L1 speakers can anticipate phonological form under optimal conditions, although such effects are small, inconsistent, and short-lived, amounting on average to an approximately 5% increase in looks and lasting roughly 600 ms when observed. In the present experiment, the increased speech rate (3.5 vs. 2.5 syllables/second in Ito and Husband) was expected to enhance temporal alignment between linguistic input and eye-movement behavior (Fernandez et al., 2020), potentially creating more favorable conditions for detecting phonological prediction. Nevertheless, given the limited and inconsistent evidence for phonological prediction in the literature, I expected at most weak and transient phonological competitor effects in L1 typical readers, consistent with the meta-analytic estimates reported by Ito (2024). For L2 speakers, evidence for phonological prediction is even more limited. Studies using the English *a/an* determiner alternation have found that proficiency modulates phonological anticipation, with higher-proficiency L2 speakers showing earlier and stronger effects (Connell et al., 2021). However, research examining whether L2 speakers pre-activate the phonological form of predictable targets strongly enough for competitors to be activated has yielded largely null results (Ito et al., 2018). Therefore, I predicted minimal or absent phonological competitor effects in L2 speakers. For adults with dyslexia, I predicted no phonological competitor effects. This prediction was motivated by consistent evidence of phonological processing difficulties in dyslexia, including reduced sensitivity to phonological information during spoken word recognition (Desroches et al., 2006; Schwarz et al., 2024) and impaired phonological awareness across the lifespan (Snowling et al., 2020).

Individual differences mechanisms. If any observed group differences in prediction reflected limitations in processing efficiency, prediction timing should correlate with indices of lexical access and automatization, in particular rapid automatized naming performance and reading history (ARHQ) scores. By contrast, if group differences or individual variability in anticipatory processing reflect cumulative literacy experience rather than processing efficiency per se, stronger relationships with vocabulary knowledge (PPVT) and print exposure (ART) measures would be expected, although such correlations may be weaker or less systematic given the highly educated, university-level sample. Alternatively, the absence of systematic correlations between anticipatory measures and individual-differences variables would suggest that group-level differences reflect broader qualitative differences in predictive engagement rather than modulation by individual cognitive or experiential factors. For L2 speakers, analogous expectations applied to measures of language experience and proficiency derived from the LEAP-Q, with greater exposure and higher proficiency potentially supporting earlier or stronger anticipatory effects (Chambers & Cooke, 2009; Tagliani et al., 2025).

4.2 Data Analysis and Results

4.2.1 Language Experience and Individual Differences

Table 4.2: *Descriptive statistics and group comparisons for individual difference measures*

	Group Means (SD)			Group Comparisons		
	L1 Typical readers	Dyslexia	L2	L1 vs Dys	L1 vs L2	Dys vs L2
RAN Letters Rate	0.37 (0.07)	0.48 (0.09)	0.37 (0.10)	$p < .001$	$p = .88$	$p < .001$
RAN Digits Rate	0.35 (0.07)	0.41 (0.07)	0.35 (0.09)	$p = .001$	$p = .98$	$p = .008$
ARHQ Sum	35.95 (8.55)	54.45 (6.48)	33.67 (7.22)	$p < .001$	$p = .26$	$p < .001$
ADHD Hyperactivity-Impulsivity	23.85 (6.35)	28.34 (7.36)	26.17 (6.95)	$p = .011$	$p = .19$	$p = .27$
ADHD Inattention	27.23 (5.34)	34.07 (6.57)	28.62 (5.25)	$p < .001$	$p = .31$	$p = .002$
ADHD Overall Mean	51.15 (10.52)	62.55 (12.70)	54.79 (11.06)	$p < .001$	$p = .20$	$p = .021$
PPVT Standard Score	103.59 (9.67)	101.62 (13.33)	103.58 (10.68)	$p = .50$	$p = 1.0$	$p = .55$
ART Score	8.49 (4.97)	7.59 (5.48)	11.00 (9.14)	$p = .49$	$p = .23$	$p = .12$
Stroop Difference	70.61 (119.49)	155.11 (232.94)	46.72 (109.37)	$p = .08$	$p = .42$	$p = .031$

Note. Values show Mean (SD). RAN = Rapid Automatized Naming; ARHQ = Adult Reading History Questionnaire; ADHD = Attention Deficit Hyperactivity Disorder; PPVT = Peabody Picture Vocabulary Test; ART = Author Recognition Test.

Group characterization. Table 4.2 presents descriptive statistics and group comparisons for individual difference measures. Participants with dyslexia demonstrated the expected pattern of reading, processing speed, and attention-related difficulties. Compared to the typical reader control group, participants with dyslexia showed significantly slower rapid automatized naming for both letters and digits ($p \leq 0.001$ for all comparisons), higher scores on the ARHQ indicating greater lifetime reading difficulties ($p < 0.001$), and elevated ADHD symptoms (overall mean: $p < 0.001$). In contrast, the dyslexia group performed comparably to typical readers on receptive vocabulary (PPVT), print experience (ART), and executive control as indexed by the Stroop task. Although participants with dyslexia showed numerically larger Stroop interference effects than L1 typical readers, this difference did not reach conventional levels of statistical significance in the full sample ($p = 0.08$). Inspection of the data revealed one L1 typical reader with an extreme Stroop Difference score, while showing no outlying values on any other individual-differences measure. Given the absence of a principled reason for exclusion, this participant was retained in all analyses. A follow-up analysis excluding this individual yielded a significant group difference between L1 typical readers and participants with dyslexia ($p = 0.035$). Taken together, this pattern suggests that their difficulties were specific to core dyslexia-related processing abilities (rapid naming, processing speed, and attention) rather than reflecting reduced literacy experience. All the participants were university students, so this profile may reflect compensated dyslexia, in which individuals develop strategies that allow them to succeed academically despite persistent underlying processing challenges. L2 speakers performed similarly to L1 controls on all measures ($p \geq 0.2$ for all comparisons), validating that any observed differences in predictive processing would reflect language-specific rather than general cognitive factors.

Moreover, expected associations among the measures were replicated in a correlation analysis, providing validation for our cognitive assessment battery. Consistent with expectations, participants with greater reading difficulties (higher ARHQ scores) showed significantly less print exposure (ART: r = -0.350, $p < 0.001$) and lower vocabulary scores (PPVT: r = -0.276, $p < 0.01$). Additionally, the strong correlation between reading history difficulties and ADHD symptoms (r = 0.554, $p < 0.001$) aligns with epidemiological evidence of $\sim$30% comorbidity between dyslexia and attention difficulties (Pennington, 2006). Print exposure and vocabulary knowledge also correlated positively (r = 0.371, $p < 0.001$), supporting the relationship between reading experience and lexical development (Cunningham & Stanovich, 1991).

Table 4.3: *Language background characteristics from LEAP-Q assessment (Experiment 2)*

LEAP-Q Variable	L2 English Speakers
L2 (English) Language Experience:	
Age of acquisition (years)	6.4 (2.4)
Daily exposure to English (%)	71.7 (15.3)
Daily exposure to L1 (%)	25.8 (11.1)
English academic/professional exposure (years)	8.1 (5.5)
English Academic/Professional Exposure:	
Less than 1 year	8%
1–5 years	32%
More than 5 years	60%
Self-Rated L1 Proficiency:	
Average	5.5 (0.8)
Self-Rated L2 (English) Proficiency:	
Average	5.1 (0.7)

Note. Values show Mean (Standard Deviation).
English Academic/Professional exposure indicates the duration of time spent in schools and/or workplaces where English was the primary language of communication. Self-rated proficiency measured on a scale from 1 (lowest) to 6 (highest).

Language background. Table 4.3 shows that L2 participants had early English acquisition (M = 6.4 years), extensive daily exposure to English (72%), and high self-rated proficiency in both languages (L1: 5.5/6, English: 5.1/6). This L2 profile reveals some contrasts with Experiment 1 despite similar age of L2 acquisition onset (both ≈ 6). The current English L2 speakers showed reversed dominance patterns compared to the Tyrolean-Italian bilinguals, with higher L2 English (72%) than L1 daily exposure (26%) versus higher L1 Tyrolean (67%) than L2 Italian exposure (20%). This difference reflects distinct sociolinguistic contexts: while Tyrolean coexists with Italian in South Tyrol's multilingual environment (see Section 1.5), English dominates in both academic and daily life contexts in the UK. Additionally, the current L2 sample had more extensive academic/professional English exposure. Despite these different dominance patterns, both bilingual samples demonstrated similar proficiency self-ratings.

4.2.2 Data Cleaning and Analysis Approach

Data preprocessing. Eye-tracking data were recorded using an EyeLink eye-tracker and exported with fixation counts pre-aggregated per time bin: each row represented one participant × trial × time-bin observation. No prior averaging across participants or items was performed before modelling. Fixation proportions were calculated across all samples assigned to predefined interest areas, excluding off-screen looks. This approach facilitates interpretation of shifts in visual attention between areas of interest, as including off-screen looks can obscure whether increases in fixations to a given region reflect genuine attentional preferences or a general increase in on-screen looking. Trials consisting exclusively of off-screen looks were removed, as they provided no usable fixation data. To avoid including trials with too few data, a track-loss threshold of 30% was applied. Given the high temporal resolution of the eye-tracking data sampled at 500 Hz (20 ms bins), this cutoff was chosen to account for natural blinking and brief tracking interruptions while maintaining sufficient data for reliable analysis (Dink & Ferguson, 2015). Trials exceeding this threshold were excluded, resulting in the removal of 30 trials (6% of the dataset). Next, trials in which participants failed to fixate the critical object in target conditions were excluded (2% of the remaining data), as such cases likely reflected lapses in auditory comprehension, attentional disengagement, or misunderstanding of task requirements. For example, one participant initially adopted a peripheral viewing strategy, avoiding direct fixations while still selecting targets correctly; after clarification of task instructions, this participant performed normally in the subsequent block.

At the participant level, individuals contributing fewer than 59 trials (determined by -1.5 SD below the mean number of completed trials) were excluded to ensure adequate statistical power. This resulted in the removal of one participant from the dyslexia group, who contributed only 49 out of 116 possible trials. Finally, participant-level outliers on the individual differences battery were identified using within-group z-scores. Participants showing extreme values with z-scores exceeding ± 2.5 on multiple standardized cognitive measures were flagged for review. One control participant met this criterion, exhibiting extreme scores on three key measures: rapid automatized naming (RAN) digits rate (z = 2.54), RAN letters rate (z = 2.70), and Stroop difference (z = -4.38), indicating performance that was substantially atypical compared to their group. This participant was therefore excluded to prevent disproportionate influence on group-level analyses. After preprocessing and exclusions, the final dataset comprised 93 participants, each contributing an average of 105.36 trials (SD = 15.61), for a total of 116 distinct experimental trials.

Analysis approach. To visually inspect the data, Figure 4.3 displays the time-course of fixation proportions to critical objects across all experimental conditions for each participant group in both predictable and unpredictable items. Visual inspection already reveals clear anticipatory effects emerging well before target word onset and sustained effects continuing into the post-target period in predictable items. In unpredictable contexts, on the other hand, all conditions show similar, low fixation proportions during the pre-target period, with sharp increases only after target word onset.

To statistically compare these patterns, two complementary analytical approaches were employed. First, Generalized Additive Mixed Models (GAMMs) were implemented using the *bam()* function from the `mgcv` package in R (Wood, 2017) to test whether the smooths (time courses of fixation proportions) differ between groups or conditions across the two primary time windows: an anticipatory window (-3400 to 200 ms relative to target onset) capturing prediction-based processing before acoustic information becomes available, and a resolution window (200 to 800 ms) capturing lexical access and integration processes following target word onset. Post-hoc smooth comparisons were carried out using functions from the `itsadug` package (e.g., `plot_diff()`, `get_smooths_difference()`) combined with cluster-based permutation tests. These analyses provided the GAMM-equivalent of pairwise contrasts, allowing me to identify the specific time spans where fixation trajectories significantly diverged. This procedure

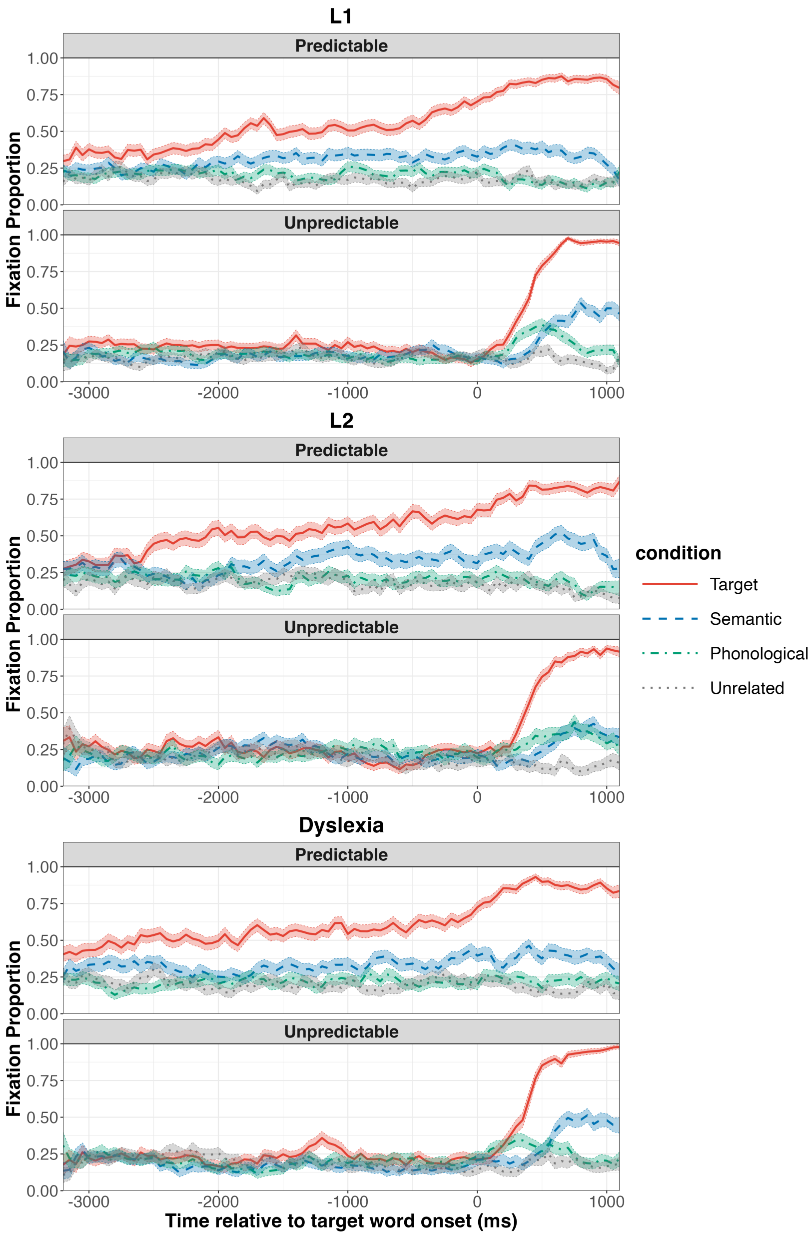


Figure 4.3: Mean fixation proportions on target, semantic competitor, phonological competitor, and unrelated objects in the L1 (top graph), L2 (middle graph), and Dyslexia (bottom graph) groups. Each panel shows data for one participant group, with the upper portion displaying predictable sentence contexts and the lower portion showing unpredictable contexts ("If there is one, click on the picture of the...").

was applied systematically to all between- and within-group comparisons. Second, divergence point analysis (DPA) was used to estimate the temporal onset of condition differences (see Section 2.2.5 for methodological details). Together, these approaches enabled a detailed examination of group- and condition-specific patterns of anticipatory processing.

To find the best-fit model, my selection process was similar to Experiment 1 and involved two key comparisons. First, I constructed models with varying random effects structures: a simple model with participant- and item-specific random smooths over time that did not vary by experimental condition, and a more complex model that allowed participant-specific smooths to vary by experimental condition. This choice was motivated by the expectation that individual differences in predictive processing may manifest differently across conditions and participant groups. Item-level variability was modeled using item-specific factor-smooths over time (`s(time, item, bs = ‘‘fs’’)`); item-by-condition smooths were also considered but were not retained due to overfitting and lack of improvement in model fit. Second, models with and without the AR(1) autocorrelation (ρ) parameter were compared to determine whether including the AR(1) correlation structure significantly improved model fit. The condition $\times$ group interaction was encoded in a single ordered factor variable (`cg`) with treatment (dummy) coding. The reference level was `Unrelated.L1` (the unrelated condition in the L1 typical-reader group) such that all parametric coefficients reflect contrasts against mean fixation in that baseline cell, following the same coding rationale as in Experiment 1a (see Section 2.2.4). As in Experiment 1, autocorrelation control was important given the dense temporal sampling of eye-tracking data, where fixations at adjacent time points are highly correlated due to the rapid nature of eye movements. To enable autocorrelation modeling in GAMMs and to handle zero and one proportions, the dependent variable was the empirical logit (*Elog*) of fixation proportions, computed directly from the binned sample counts as:

$$Elog = \log\left(\frac{c + 0.5}{n - c + 0.5}\right)$$

where c is the number of gaze samples falling on the critical object within a given time bin and n is the total number of valid samples in that bin. The correction factor of 0.5 was added to both numerator and denominator to handle cases where $c = 0$ or $c{=}n$ (Barr, 2008).

The final models for both time windows included:

- fixed smooths of time and condition-by-group interactions,
- condition-specific smooths of time to capture non-linear trajectories,
- participant-specific smooths varying by condition to account for individual variability,
- item-specific smooths of time,
- and an AR(1) autocorrelation (ρ) parameter to control for temporal dependencies.

4.2.3 Results: Anticipatory Window

First, eye movement data in the anticipatory window (-3400 to 200 ms) were analyzed using GAMMs. Model comparisons demonstrated that a complex random effects structure, which allowed participant-specific slopes to vary by condition, provided a better fit than the simpler model (ΔfREML = –890.34, Δdf = 20, $p < .0001$, AIC difference = 3234.41).This indicates that individuals exhibited condition-specific anticipatory patterns, justifying the inclusion of by-condition participant smooths in the final model. See the full model summary in Appendix F.1.

Within-group anticipatory effects. All three participant groups demonstrated clear anticipatory fixations to target objects consistent with predictive processing, with fixation proportions beginning to rise approximately 3000 ms before target word onset and sustaining throughout the anticipatory window (see Figures 4.4, 4.5).

According to the GAMM analysis, in the L1 group, fixations to the target started to diverge from unrelated items from approximately -2927 ms (effect size = 0.705), while semantic competitors also elicited anticipatory fixations

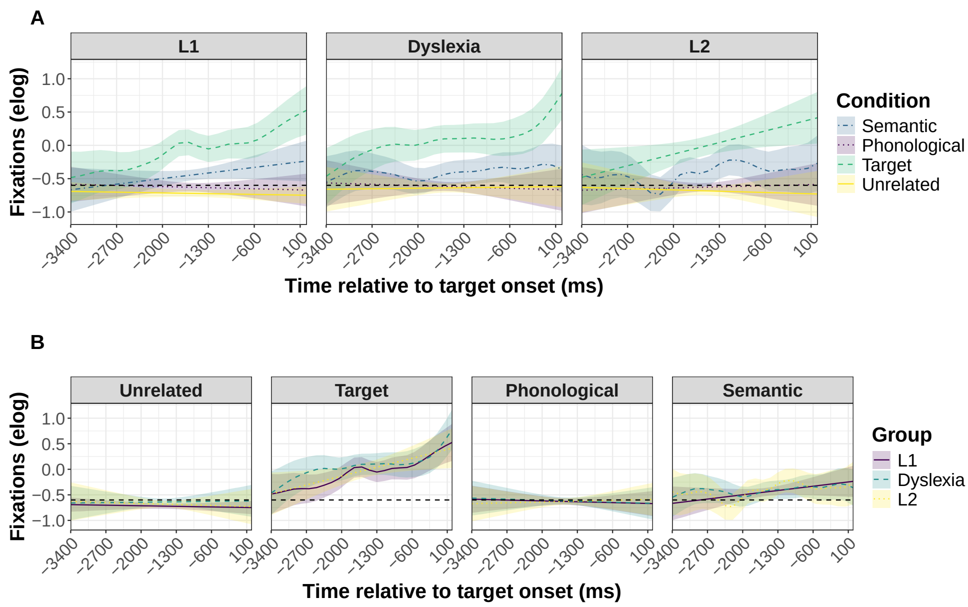
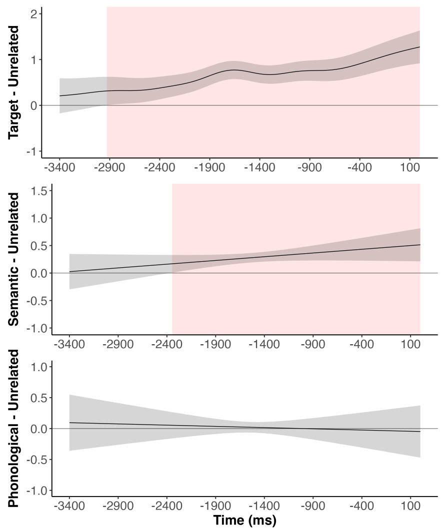


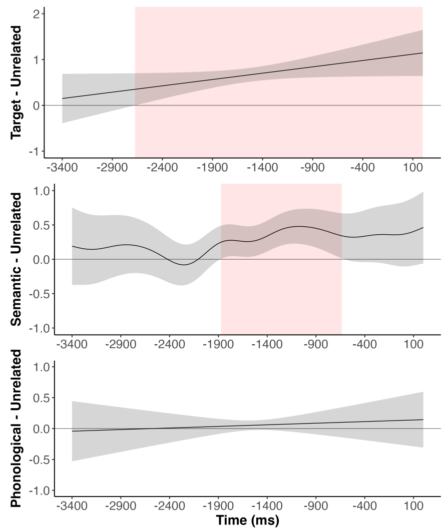
Figure 4.4: GAMM model estimates of looks to the critical object over time. (A) Comparison by condition for each group. (B) Comparison by group for each condition. The x-axis represents time relative to the target word onset (0 ms), and the y-axis shows the proportion of looks directed to the critical word transformed into empirical logit.

relative to unrelated distractors starting from around -2345 ms, lasting until the end of the time window (effect size = 0.341). The dyslexia group showed anticipatory fixations to the target emerging from around -3036 ms (effect size = 0.746), and slightly weaker semantic competitor activation from -1618 to -600 ms (effect size = 0.255). Similarly, L2 participants anticipated targets from -2673 ms (effect size = 0.748) and demonstrated semantic competitor activation from -1873 to -636 ms (effect size = 0.367). Thus, each population demonstrated predictive eye movements, although the timing and strength of semantic competitor activation varied across groups.

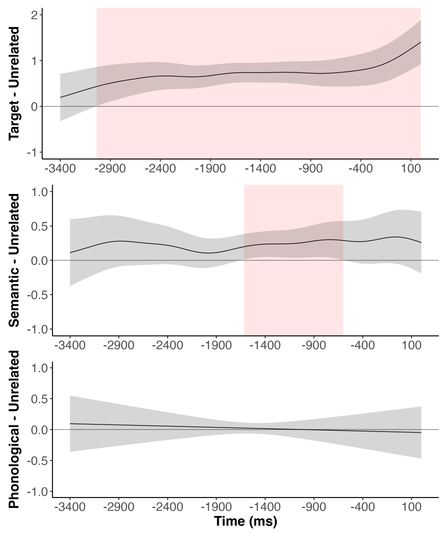
In contrast to significant semantic effects, no evidence emerged for phonological competitor activation during the anticipatory window across any participant group. A simulation-based power analysis was conducted to assess statistical sensitivity to detect a phonological competitor effect of the magnitude identified in the meta-analysis by Ito (2024), following



(a) L1 English typical readers



(b) L2 English speakers



(c) Participants with dyslexia

Figure 4.5: Panels show differences in empirical logit-transformed fixation proportions over time between the target and unrelated conditions (upper graph in each panel), semantic and unrelated conditions (middle graph), and phonological and unrelated conditions (lower graph in each panel) for (a) L1 English typical readers, (b) L2 English speakers, and (c) participants with dyslexia. The x-axis represents time relative to target word onset (0 ms).

the same approach as Experiment 1a (see Section 2.2.5). The effect size was operationalized as a 5% absolute increase in fixation proportion from the empirical baseline. Two power estimates were computed. Using the full sample (n = 93), estimated power was 86.9%, exceeding the conventional 80% threshold and indicating adequate sensitivity at the level of the overall design. However, this estimate assumes a homogeneous effect of the same magnitude across all three participant groups, whereas I predicted that phonological prediction might be attenuated or absent in L2 speakers and adults with dyslexia (see Section 4.1.5). Restricting the analysis to the L1 typical reader group alone (n = 39), where phonological prediction effects would most plausibly emerge, yielded an estimated power of 55.7%, below the conventional threshold. While the design was not negligible in power, the estimated sensitivity suggests that the phonological competitor effect would be detected with only moderate probability, and the null result should therefore be interpreted with appropriate caution. Nonetheless, the Experiment 2 estimates represent a substantial improvement over Experiment 1a (86.9% and 55.7% vs. 22.8% and 11.5% for L1 readers), reflecting both the larger sample and considerably lower residual variance, the latter likely due to the higher temporal resolution of the EyeLink 1000 eye-tracker (500 Hz). The convergence of null phonological effects across both experiments therefore carries greater evidential weight than either result in isolation, with Experiment 2 providing the more informative basis for theoretical interpretation.

Between-group comparisons of anticipatory processing. Overall, the model did not reveal statistically significant between-group differences in the anticipatory window (see Figure 4.6). Importantly, the absence of significant group effects does not imply that the groups initiated prediction at the same time. GAMMs test for overall proportional differences in fixation trajectories and can detect when fixation proportions diverge between conditions within a group, but they do not test whether the timing of divergence differs significantly between groups. In fact, visual inspection suggested that dyslexic participants displayed slightly higher baseline looks to both predictive and unrelated items, therefore reducing the relative contrast between conditions. As a consequence, group-level differences in fixation proportions may not be reflected in GAMM contrasts. To capture these temporal aspects, DPA was used.

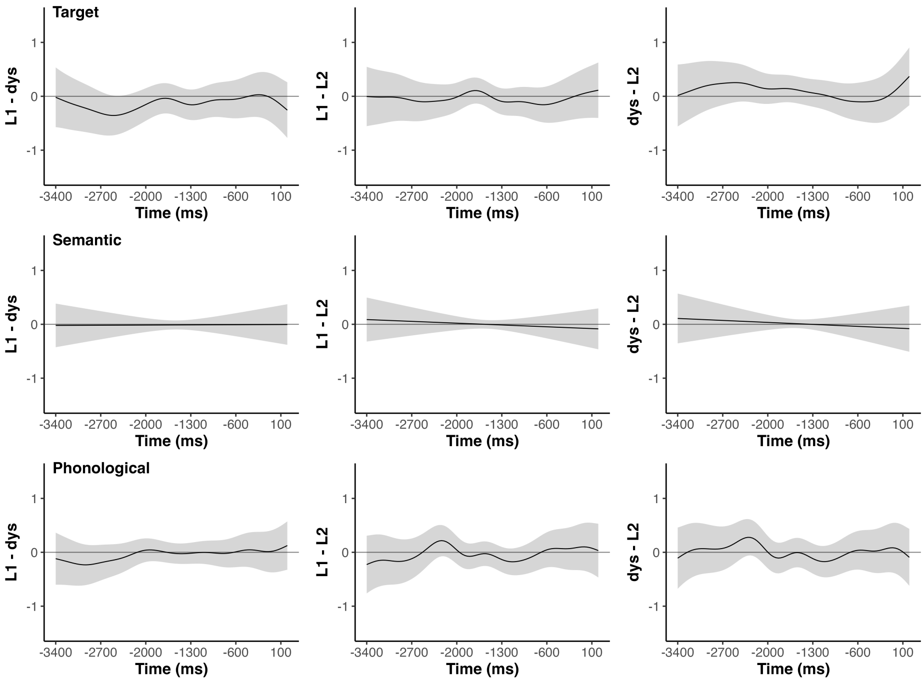
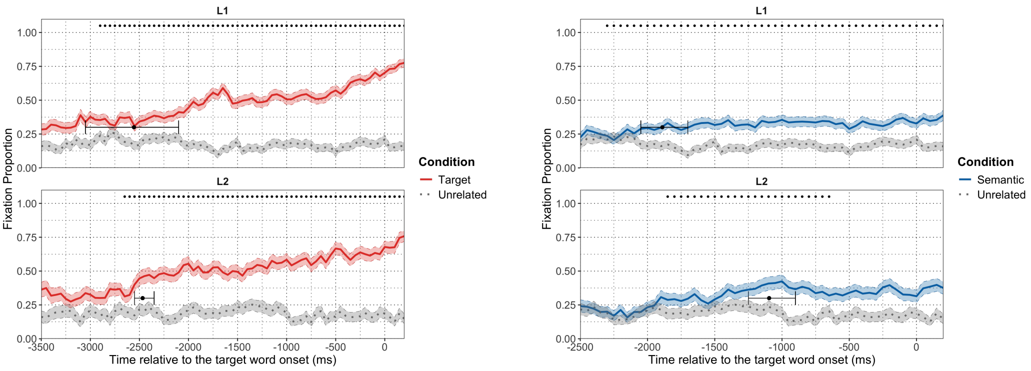


Figure 4.6: Panels show differences in fixation proportions over time between groups in the target (upper panel), semantic (middle panel), and phonological (lower panel) conditions across the anticipatory time window for L1 typical readers vs. participants with dyslexia (leftmost panels), L1 vs. L2 participants (middle panels), and dyslexics vs. L2 speakers (rightmost panels).

When comparing **L1 and L2 speakers**, divergence point analysis did not reveal significant timing differences in target divergence onsets in the anticipatory window (see Figure 4.7a). Both groups initiated anticipatory looks at comparable time onsets with L1 speakers' mean onset of -2559.7 ms (95% CI: [-3050, -2100]), and L2 speakers' mean onset at -2467.55 ms (95% CI: [-2550, -2350]). The estimated difference of just 42 ms was non-significant (95% CI: [-450, 650], $p = 0.2$).

For semantic competitors, however, a different pattern emerged (see Figure 4.7b). L1 participants diverged earlier (at -1892.95 ms; 95% CI: [-2050, -1700]) than L2 speakers (at -1097.6 ms; 95% CI: [-1250, -900]), a delay of roughly 745 ms for the L2 group. Although this difference was



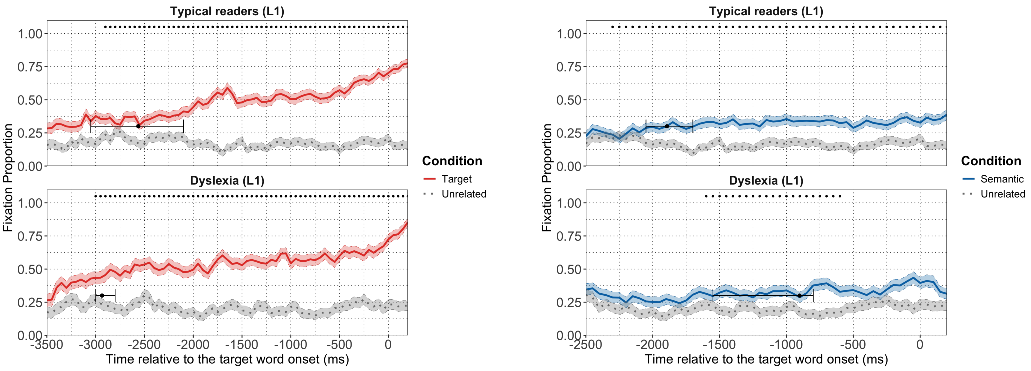
(a) (b)

Figure 4.7: Divergence point analysis of mean fixation proportions to the critical word in a) target vs. unrelated conditions, b) semantic vs. unrelated conditions for L1 and L2 groups.

visually striking and supported by non-overlapping confidence intervals, the permutation-based test did not reach conventional significance (95% CI: [550, 1050], p = 0.06). This discrepancy likely reflects the high variability within the L2 group, which broadened the null distribution and increased the probability of observing large apparent differences by chance. Thus, while not statistically reliable, the pattern shows a clear trend toward delayed semantic anticipation among L2 speakers.

Comparison between **individuals with dyslexia and typical readers (controls)** did not reveal any statistically significant differences in target prediction between the two groups (see Figure 4.8a). Both groups initiated target prediction well before word onset: typical readers at a mean of -2559.7 ms (95% CI: [-3050, -2100]), and dyslexics at -2935.45 ms (95% CI: [-3049, -2850]). The small difference in divergence timing of a mean 421 ms was not significant (95% CI: [-360, 80], p = 0.55).

In contrast, clear group differences emerged for semantic competitors (see Figure 4.8b). While typical reader controls directed anticipatory fixations toward semantic competitors from around -1892.95 ms (95% CI: [-2050, -1700]), participants with dyslexia did not show reliable divergence until a mean of -902.8 ms (95% CI: [-1550, -800]), yielding a significant delay of approximately 939 ms (95% CI: [300, 1250], $p < 0.01$). This finding



(a) (b)

Figure 4.8: Divergence point analysis of mean fixation proportions to the critical word in a) target vs. unrelated conditions, b) semantic vs. unrelated conditions for typical readers and participants with dyslexia.

suggests that although dyslexic readers anticipate predictable targets, they require additional time to engage in broader semantic anticipatory processing compared to typical readers.

Finally, the comparison between **L2 and dyslexic** participants revealed a broadly similar profile of anticipatory behavior. For target fixations, participants with dyslexia actually showed a numerical advantage, diverging approximately 423 ms earlier than L2 participants, though this difference did not achieve significance (95% CI: [300, 650], $p = 0.06$). For semantic competitors, divergence points were closely aligned (L2 = –1097 ms; dyslexia = –902.8 ms), and the small group difference of 245 ms was non-significant (95% CI: [-400, 550], $p = 0.9$)[1].

Taken together, these comparisons indicate that dyslexic and L2 participants engage in contextual prediction at levels comparable to L1 speakers. However, both groups showed evidence of delayed initiation of semantic anticipatory processing. For dyslexic readers, this delay was robust and statistically reliable relative to typical readers, whereas for L2 participants, the pattern trended in the same direction but did not reach

[1] As the corresponding trajectory plots for both groups were already shown in the earlier between-group contrasts (Figures 4.7, 4.8), additional panels are not reproduced here to avoid redundancy; only the divergence estimates are reported.

significance, suggesting a possible but unconfirmed delay in anticipatory processing.

Competitor type analysis: CP+Sem vs. CP-only. Substantial within-group variability in the initial semantic competitor effects, particularly pronounced in the dyslexia group, suggested that the heterogeneity was not merely random noise but might reflect systematic differences in the nature of semantic relationships within the experimental stimuli. This observation prompted a detailed examination of the semantic competitor items to identify potential sources of heterogeneity that could account for the observed patterns. Upon the review of all semantic competitor items, a meaningful pattern emerged. Some semantic competitors appeared to be exclusively related to the highly predictable target word (e.g., in "The restaurant is always busy, so Leo will book a *table* for the whole family", the semantic competitor *drawers* relates solely to the target *table*). Others, by contrast, were related not only to the target but also to additional content words in the preceding sentence context (e.g., in "That dog looks so happy, wagging its *tail* as it walks along", the semantic competitor *paw* is related both to the target word *tail* and the contextually preceding word *dog*).

Therefore, this distinction suggested that semantic competitors might be activated through different mechanisms. Items linked only to the target are likely driven by *target-based activation*, becoming active once the target itself is predicted. Items linked to both the target and context words may benefit from *context-driven activation*, in which spreading activation from multiple contextually related words facilitates earlier and stronger competitor engagement. Based on this rationale, semantic competitors were divided into:

- **CP-only:** items semantically related to high Cloze Probability (CP) target alone;
- **CP+Sem:** items semantically related to both the CP target and additional contextual words.

To validate this theoretical distinction and provide empirical support for the post-hoc classification, a semantic relatedness norming study was conducted with a convenience sample of 33 native English speakers who had not participated in the main eye-tracking experiment. There is no established minimum sample size for semantic relatedness norming. Although norming procedures differ in the type of judgments elicited (e.g., cloze probabilities, semantic fit, plausibility, lexical associations), psycholinguistic research typically reports samples between 15 and 30 participants (e.g., Aveni et al., 2023; Brouwer et al., 2019; DeLong et al., 2005; Mishra et al., 2012; Staub & Clifton, 2006). Participants completed an online word association task in which they viewed identical picture representations of semantic competitor objects used in the eye-tracking experiment. This methodological choice to use pictures instead of written words ensured that semantic relatedness judgments reflected responses to the actual visual stimuli that participants encountered during predictive processing, avoiding potential discrepancies between written word associations and visual object recognition.

For each semantic competitor image, participants were presented with four word options plus a separate "none of the above" option. The four words comprised (1) the target word from the corresponding sentence, (2) content words from the sentence context preceding the target (including nouns, adjectives, and verbs that carried semantic weight), (3) unrelated words included to complete the four-option format when necessary. Participants first read task instructions stating that they would see an image followed by a list of words and should select all words they considered related to the image. They were explicitly instructed that relationships could encompass meaning, function, typical situations, or common connections and that there were no correct or incorrect responses (Appendix I shows the instruction screen). For example, for the sentence "That dog looks so happy, wagging its *tail* as it walks along," participants were presented with a picture of a *paw* together with content words *dog* and *to wag*, an unrelated word *bottle*, and the target word *tail* (see Figure 4.9).

A Context Relationship Score (CRS) was developed specifically for

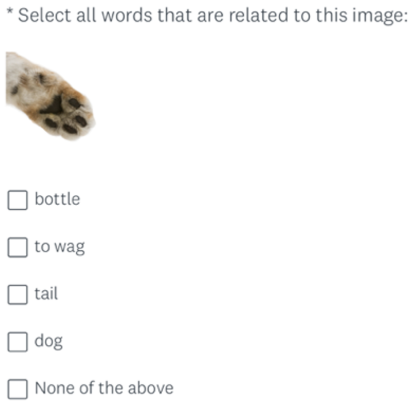


Figure 4.9: Example of a norming task showing a semantic competitor image of a *paw* with word association options for the sentence: "That dog looks so happy, wagging its *tail* as it walks along."

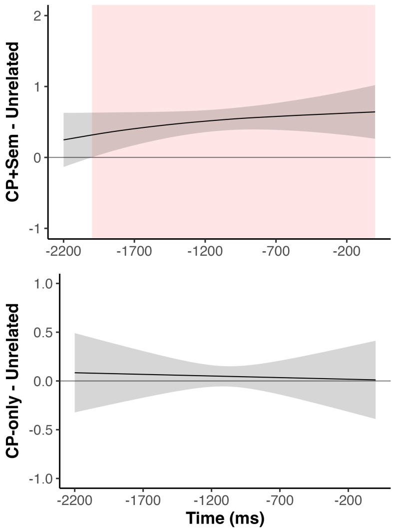
this study to quantify the strength of association between each semantic competitor and the sentence context (excluding the target word). Because participants could select more than one contextual word for each item, and because the target word was always present among the response options, a simple proportion measure could not accurately capture the overall degree of contextual activation. The CRS therefore summarises both how many contextual elements were judged to be related to the competitor (Context Sum) and how consistently these relationships were endorsed across participants (adjustment for None%). The score was defined as:

$$\text{CRS} = \text{Context Sum} \times \left(1 - \frac{\text{None\%}}{100}\right)$$

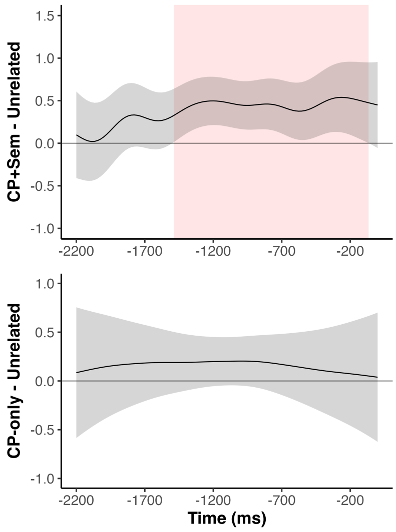
where *Context Sum* represents the cumulative selection percentages for all contextual words (excluding the target) identified as related to the competitor, and *None%* represents the proportion of participants who identified no relationship between the competitor and context words. Because participants could select multiple contextual words for a given competitor, maximum CRS values could exceed 100 when items were judged to be strongly related to multiple contextual elements. Inspection of the CRS distribution showed a clear gap between items with minimal contextual activation (CRS values between 0 and 10) and the next highest item (CRS = 17), with no values falling in the intermediate range. To obtain a categorical distinction aligned with this empirical structure, a threshold was placed within this empty interval. Items with $CRS < 15\%$ were classified as CP-only (indicating relationship primarily with the contextually predicted target word), and items with $CRS > 15\%$ as CP+Sem (indicating associations not only to the target but also to one or more contextual words in the sentence). The final classification resulted in 20 CP-only items (CRS range: 0-10, M = 2.3, SD = 2.6) and 34 CP+Sem items (CRS range: 17-191, M = 81.8, SD = 51.0). This variability within the CP+Sem group reflects differences in the degree of context-based activation, while all items exceeded the threshold that distinguished them from CP-only items.

To analyze anticipatory looks to CP+Sem and CP-only competitors, a time window starting from -2200 ms was selected. This window began slightly before the earliest divergence points observed for semantic competitors in the broader anticipatory analyses, ensuring that the models captured the initiation of context-based semantic activation (see Appendix F.2 for the model summary). The analysis revealed strong CP+Sem effects across all groups (see Figure 4.10). The L1 typical reader group showed significant CP+Sem activation from -2000 ms throughout the whole window (effect size = 0.520), while the dyslexia group exhibited activation from -1644 ms (effect size = 0.428), and L2 speakers from -1489 to -67 ms (effect size = 0.457). By contrast, no group showed significant anticipatory fixations to CP-only competitors, suggesting that semantic relatedness to the target word alone was insufficient to elicit activation.

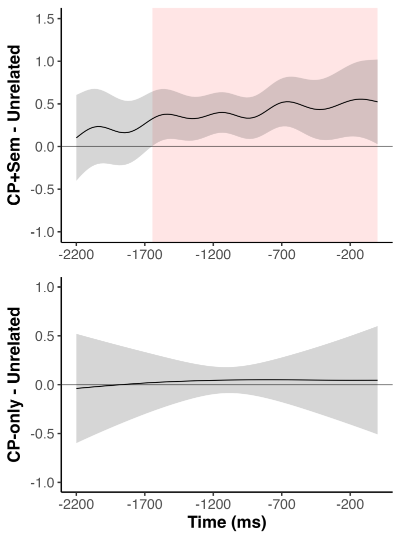
Further, divergence point analyses were performed to compare the timing of divergence in looks at CP+Sem competitors between the groups. This analysis was not conducted on CP-only items due to the non-significance of the effect.



(a) L1 English typical readers



(b) L2 English speakers



(c) Participants with dyslexia

Figure 4.10: Panels show differences in empirical logit-transformed fixation proportions over time between the CP+Sem and unrelated conditions (upper graph in each panel), and CP-only and unrelated conditions (lower graph in each panel) for (a) L1 English typical readers, (b) L2 English speakers, and (c) participants with dyslexia.

In the comparison between **typical readers and participants with dyslexia**, a clear difference emerged in divergence timing (see Figure 4.11). While the typical reader group directed anticipatory looks to CP+Sem

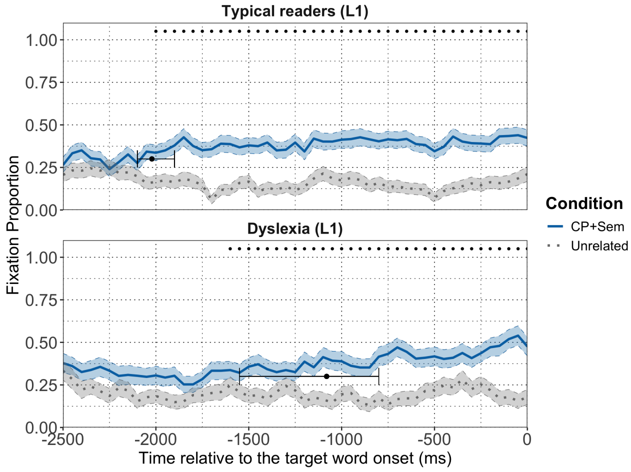


Figure 4.11: Divergence point analysis of mean fixation proportions to the critical word in CP+Sem vs. unrelated conditions for typical readers and participants with dyslexia.

competitors from approximately a mean of -2022 ms (95% CI: [-2100, -1900]), participants with dyslexia did not diverge until a mean of -1076.8 ms (95% CI: [-1550, -800]). This delay of nearly 1000 ms was statistically reliable (95% CI: [500, 1250], p = 0.02), indicating that dyslexic readers were slower to initiate semantic competitor anticipation, even when those competitors were related to both target and contextual words.

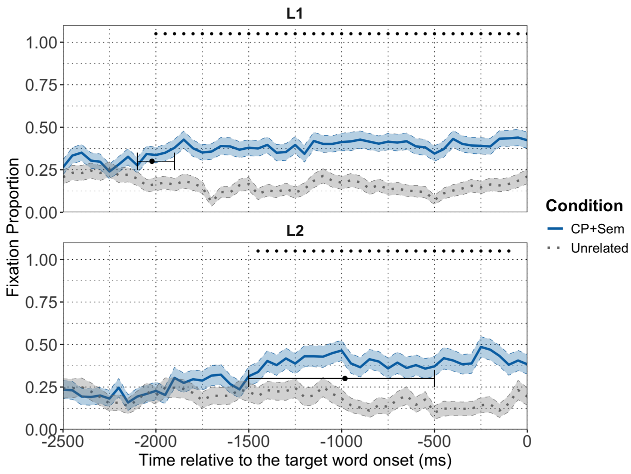


Figure 4.12: Divergence point analysis of mean fixation proportions to the critical word in CP+Sem vs. unrelated conditions for L1 and L2 speakers of English.

The contrast between **L1 and L2 participants** revealed a numerically similar pattern (see Figure 4.12). L1 participants again initiated CP+Sem fixations earlier (-2022 ms) than L2 speakers (-983 ms), corresponding to a mean delay of ∼987 ms. While ∼987 ms difference is large and visually striking in the graphs (with non-overlapping CIs), the permutation-based p-value again did not reach significance (95% CI: [550, 1900], p = 0.2). This discrepancy can be explained by the high variability in the L2 group, which broadened the null distribution generated by randomly reassigning group labels. As a result, even a difference of several hundred milliseconds could be reasonably produced by chance, leading to a non-significant p-value. Nonetheless, the pattern is likely to suggest that L2 speakers require more contextual input before engaging in semantically mediated anticipatory processes. To examine whether the substantial variability observed in the L2 group reflected systematic individual-difference patterns, I conducted an exploratory set of correlations between CP+Sem divergence points and all LEAP-Q variables. Among the LEAP-Q variables, years spent in an English-speaking country showed the largest association with CP+Sem divergence point (r = 0.527, p = 0.008), with all remaining variables producing negligible-to-small effects ($|r| < 0.30$; see Appendix H, Table H.1). Cook's distance diagnostics identified one influential observation for this association (Cook's D = 0.334, compared to 0.167 threshold): a participant, who combined the highest immersion time in the sample (18 years) with a near-zero divergence point (–65 ms, compared to a group mean of –1474 ms). Removing this participant reduced the correlation to r = 0.398 (p = 0.060), and LOO correlations ranged from r = 0.398 to r = 0.648, indicating that the estimate is sensitive to this single atypical profile. Furthermore, the direction of the association is difficult to interpret: a positive coefficient indicates that longer immersion is associated with later rather than earlier semantic anticipation. Given the instability of the estimate and its lack of significance upon removal of one influential observation (r = 0.398, p = 0.060), this association is not treated as a reliable individual-difference effect. The remaining LEAP-Q variables yielded no associations of meaningful magnitude, and no coherent pattern emerged across predictors. The variability in L2 CP+Sem anticipation, therefore, most plausibly reflects unsystematic between-participant dispersion rather than any specific underlying experiential or proficiency factor.

When comparing **L2 speakers and participants with dyslexia**, divergence points were virtually identical. Dyslexic readers diverged at a mean of -1077 ms, while L2 learners diverged at a mean of -983 ms, and the difference of just 33 ms was not significant (p = 0.9).

Taken together, these results demonstrate that semantic activation emerged only when competitor items were semantically integrated with the broader sentence context and that anticipation of a highly predictable target word alone did not produce spreading activation to its semantic associates. Individuals with dyslexia showed a consistent and statistically significant delay in initiating semantic anticipation, even though their pre-activation of highly predictable targets was intact. In contrast, L2 participants displayed a numerically similar delay relative to L1 controls, but this effect did not reach statistical significance.

Exploratory analyses: prediction and individual differences. To assess whether individual cognitive and reading-related skills modulated predictive processing, correlation analyses were conducted between participants' divergence point measures and the cognitive assessment battery (RAN, ARHQ, ADHD, PPVT, ART, Stroop). For target and semantic divergence points, analyses were conducted across the full sample; these results are reported in Appendix H, Tables H.2, H.3, and revealed no significant associations with any individual-difference measure (all $p > 0.05$). The semantic divergence point table is additionally limited in interpretive value given the subsequent finding that the semantic competitor condition comprised two qualitatively distinct subsets of items (CP+Sem and CP-only), only one of which elicited anticipatory activation (see Section 4.2.3). Accordingly, the aggregate semantic divergence point conflates items that behaved very differently and is not discussed further.

For the theoretically critical CP+Sem divergence point, correlations were first conducted across the full sample and are reported in Appendix H,

Table H.4. Although no associations reached significance in the full-sample analysis, a meaningful pattern emerged when examining effect sizes across measure types. Phonological processing speed (RAN Letters: r=-0.082; RAN Digits: r=-0.044), reading history (ARHQ: r=-0.056), and attentional control (ADHD: r=-0.044) showed negligible associations with CP+Sem divergence timing. By contrast, receptive vocabulary (PPVT: r=-0.184) and print exposure (ART: r=-0.132) showed effects of small-to-medium magnitude, suggesting that experience-based linguistic measures may relate more closely to semantic prediction efficiency than phonological or attentional measures. This differential pattern warranted further investigation at the group level. Since the three groups were selected on the basis of differing cognitive and linguistic profiles, and were therefore expected to differ systematically on the individual-difference measures themselves, group-specific correlation analyses were conducted separately for each group to provide a more interpretively transparent picture of within-group associations (see Tables 4.4, 4.5, 4.6).

Within the L1 typical reader group, no significant associations were observed between CP+Sem divergence point and any individual-difference measure (all $p > 0.19$, all $|r| < 0.2$; see Table 4.4). Effect sizes were uniformly small across all measures, suggesting that within a neurotypical L1 population, individual variation in these cognitive and linguistic skills does not reliably predict the timing of semantic anticipatory processing. Within the dyslexia group, a significant negative correlation emerged between CP+Sem divergence point and PPVT standard score ($r = -0.413$, $p = 0.026$; see Table 4.5), indicating that dyslexic participants with higher receptive vocabulary knowledge tended to initiate CP+Sem competitor anticipation earlier. A marginal association was also observed for Stroop interference score ($r = 0.346$, $p = 0.066$), suggesting a possible tendency for participants with greater interference costs to show later CP+Sem divergence, though this did not reach conventional significance. No other measure reached significance in the dyslexia group (all $p > 0.30$). Within the L2 group, a significant negative correlation emerged between CP+Sem divergence point and ART score ($r = -0.435$, $p = 0.034$; see Table 4.6), indicating that L2 participants with greater print exposure tended to show earlier CP+Sem anticipation. No other measure reached significance in the L2 group (all $p > 0.47$, all $|r| < 0.16$).

Table 4.4: *Correlations between CP+Sem divergence point and individual-difference measures in the L1 typical reader group.*

Variable	r	p
RAN Letters (rate)	0.005	0.976
RAN Digits (rate)	0.005	0.977
ARHQ total score	−0.058	0.724
ADHD overall score	0.140	0.396
PPVT standard score	−0.036	0.828
ART score	0.212	0.194
Stroop interference score	−0.166	0.313

Note. Pearson correlations between the CP+Sem divergence point and individual-difference measures in the L1 typical reader group ($n = 39$).

Table 4.5: *Correlations between CP+Sem divergence point and individual-difference measures in the dyslexia group.*

Variable	r	p
RAN Letters (rate)	0.198	0.302
RAN Digits (rate)	0.090	0.641
ARHQ total score	−0.038	0.847
ADHD overall score	−0.022	0.908
PPVT standard score	−0.413	0.026*
ART score	−0.176	0.361
Stroop interference score	0.346	0.066

Note. Pearson correlations between the CP+Sem divergence point and individual-difference measures in the dyslexia group ($n = 29$).

Taken together, these results reveal a theoretically coherent pattern across groups: phonological processing speed, reading history, and attentional measures showed consistently negligible associations with CP+Sem anticipation timing across all three populations, whereas experience-based measures — receptive vocabulary (PPVT) in the dyslexia group and print exposure (ART) in the L2 group — emerged as the strongest individual-level predictors of semantic prediction efficiency. In the dyslexia group, PPVT but not ART reached significance, which may reflect the particular relevance of vocabulary knowledge as a compensatory resource in this population.

Table 4.6: *Correlations between CP+Sem divergence point and individual-difference measures in the L2 group.*

Variable	r	p
RAN Letters (rate)	−0.148	0.490
RAN Digits (rate)	−0.155	0.470
ARHQ total score	0.028	0.898
ADHD overall score	−0.150	0.483
PPVT standard score	−0.020	0.927
ART score	−0.435	0.034*
Stroop interference score	−0.132	0.540

Note. Pearson correlations between the CP+Sem divergence point and individual-difference measures in the L2 group ($n = 25$).

Dyslexic readers with richer receptive vocabularies may draw on stronger lexical-semantic representations, facilitating earlier semantic anticipation. In the L2 group, by contrast, print exposure (ART) may provide a more sensitive index of the strength and accessibility of lexical representations, as it reflects cumulative experience with the language in context. Both findings should nonetheless be interpreted cautiously, as PPVT and ART accounted for approximately 17% ($r^2 \approx 0.171$) and 19% ($r^2 \approx 0.189$) of variance in CP+Sem divergence timing within their respective groups, leaving the majority of individual variability unexplained. The absence of significant correlations in the L1 typical reader group may reflect several possible factors. First, within-group variance in PPVT and ART was indeed more restricted in the L1 typical reader subsample relative to the other groups. Descriptive statistics (see Table 4.2) reveal that the standard deviation for PPVT scores in the dyslexia group was notably larger (SD=13.33) than in the L1 group (SD=9.67), and the standard deviation for ART scores in the L2 group was wider (SD=9.14) compared to the L1 group (SD=4.97). This greater dispersion in the dyslexia and L2 groups suggests that individual differences in vocabulary breadth and print exposure are more heterogeneous in these populations, providing the necessary variance range to detect correlations with anticipation timing. The more homogeneous distributions observed in the L1 group would reduce statistical power to detect such associations even if they existed. Future work could address this by recruiting L1 typical readers with a wider range of vocabulary and print exposure scores to provide sufficient variance to test whether the associations observed here in the dyslexia and L2 groups generalize to typical readers as well. Second, reduced variance in CP+Sem anticipation timing in this group may have similarly limited the detectable outcome range. Third, the association between lexical experience and semantic anticipation timing may genuinely be weaker in typical readers, not because vocabulary is irrelevant in principle, but because L1 typical readers likely cluster toward the more proficient end of these measures. Individual differences in vocabulary and print exposure may only meaningfully modulate prediction efficiency when lexical representations are less automatized, producing detectable correlations in the dyslexia and L2 groups but not in a population where processing is already well-established. Under this view, the associations observed in the dyslexia and L2 groups reflect the particular relevance of vocabulary breadth and cumulative language exposure when prediction is less efficient or less automatized — a pattern that may diminish, rather than disappear entirely, as processing becomes more entrenched.

4.2.4 Results: Resolution Window

The resolution window (200-800 ms relative to target word onset) captures lexical access and integration processes following acoustic presentation of the target word (see Figure 4.13). This analysis window encompasses the period during which participants receive bottom-up auditory information and integrate it with their contextual predictions, providing insight into how predictive processing influences word recognition and competitor activation following target onset.

Model comparison confirmed that the complex model, including both the condition × group interaction and the autocorrelation parameter (ρ), provided the best fit to the data. The complex model outperformed the corresponding null model without the interaction (ΔfREML = –11,815.19, Δdf = 33, $p < 0.0001$, AIC difference = 23,452.29), confirming substantial condition effects during the resolution period. In addition, inclusion of the autocorrelation parameter also substantially improved model fit compared

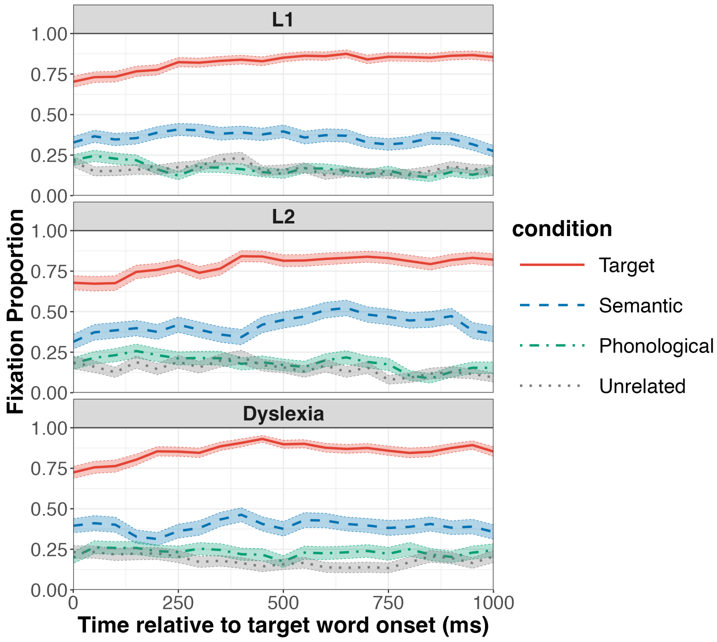


Figure 4.13: Mean fixation proportions on target, semantic competitor, phonological competitor, and unrelated objects in the L1 (top graph), L2 (middle graph), and Dyslexia (bottom graph) groups after the target onset in predictable contexts.

to the model without autocorrelation (ΔfREML = -11677.383, Δdf = 0, $p < 0.001$; AIC difference = 23,441), underscoring the importance of accounting for temporal dependencies in dense eye-tracking time series. See Appendix F.3 for model summary.

Within-group effects in predictable contexts. All groups showed sustained, high-magnitude fixations to target objects throughout the window, confirming successful target identification and integration (L1: effect size = 1.430; Dyslexia: effect size = 1.555; L2: effect size = 1.366) (see Figures 4.14, 4.15). Semantic competitor activation also persisted into the resolution window across all participant groups, though with substantially reduced magnitude compared to target effects (L1: effect size = 0.447; Dyslexia: effect size = 0.480; L2: effect size = 0.596).

Consistent with anticipatory window findings, no significant phonological competitor effects emerged during the resolution window for any participant

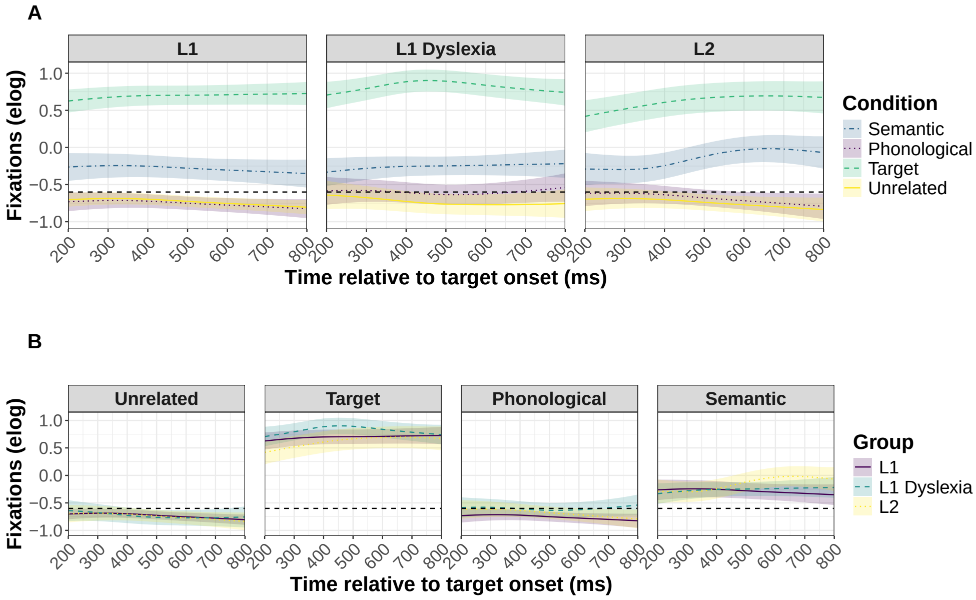
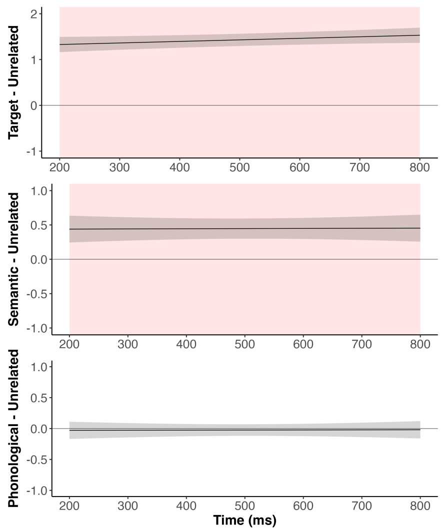
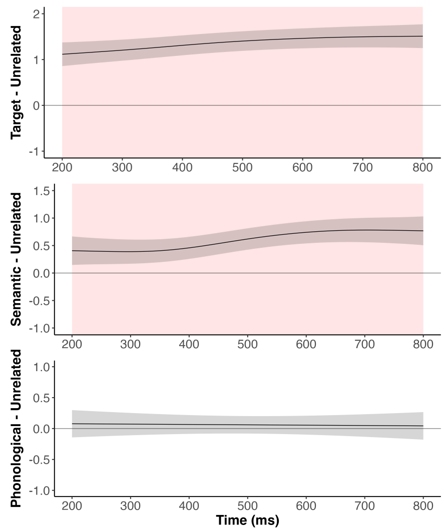


Figure 4.14: GAMM model estimates of looks to the critical object over time in the resolution window of predictable items. (A) Comparison by condition for each group. (B) Comparison by group for each condition.

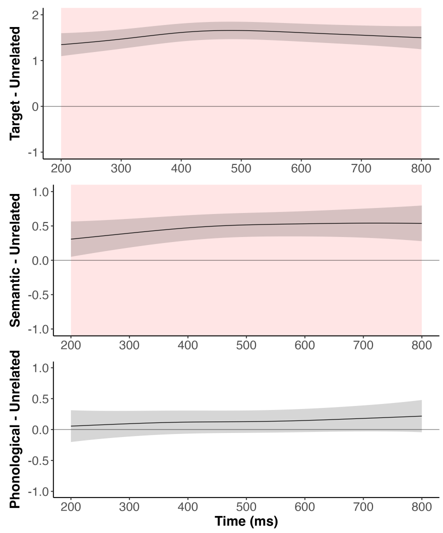
group (Figure 4.15). Similar patterns of no phonological form activation in the resolution window were observed in Experiment 1a. As argued previously, the absence of phonological competition during lexical access might indicate that strong contextual constraints effectively eliminate phonological competitor activation even when bottom-up acoustic information becomes available.



(a) L1 English typical readers



(b) L2 English speakers



(c) Participants with dyslexia

Figure 4.15: Panels show differences in empirical logit-transformed fixation proportions over time between the target and unrelated conditions (upper graph in each panel), semantic and unrelated conditions (middle graph), and phonological and unrelated conditions (lower graph in each panel) for (a) L1 English typical readers, (b) L2 English speakers, and (c) participants with dyslexia in the resolution window of predictable items.

Between-group comparisons in predictable contexts. Between-group contrasts revealed modest differences in target fixation proportions. Participants with dyslexia showed slightly fewer target looks

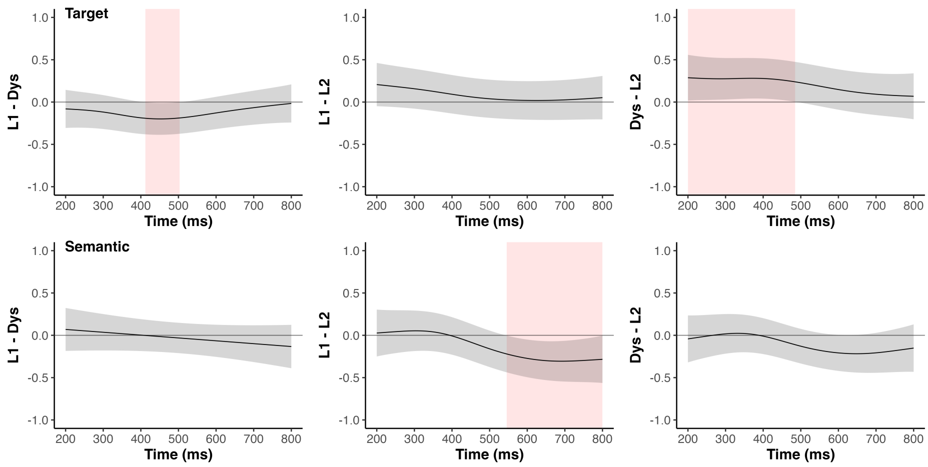
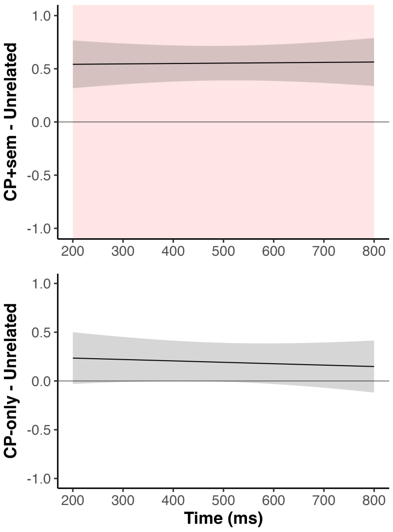


Figure 4.16: Panels show differences in fixation proportions over time between groups in the target (upper panel) and semantic (lower panel) conditions across the resolution time window for L1 English typical readers vs. participants with dyslexia (leftmost panels), L1 English vs. L2 English participants (middle panels), and dyslexics vs. L2 speakers (rightmost panels).

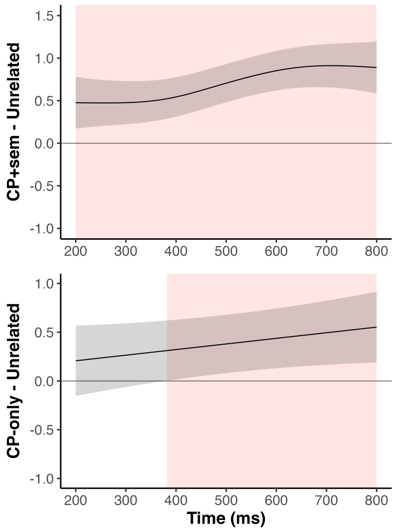
than L1 controls in a brief interval between 412–503 ms (effect size = -0.195), but they also displayed significantly more target looks than L2 participants during the early resolution phase spanning 200–485 ms (effect size = 0.274; see Figure 4.16). These effects were short-lived and weak in magnitude, suggesting that group-level differences in target identification were minimal. Overall, dyslexic readers converged on the target as efficiently as L1 typical readers and, in certain respects, more quickly than L2 participants, but the practical significance of these contrasts is limited.

By contrast, L2 participants exhibited the strongest semantic competitor activation of all groups during resolution (overall semantic effect size = 0.596), with significantly more looks to semantic competitors than L1 speakers between 545–800 ms (effect size = 0.285). Given its modest size and limited temporal extent, this effect was treated as provisional and evaluated against the unpredictable sentence analyses below, which target bottom-up lexical processing without strong contextual constraint. In these unpredictable contexts, the effect did not replicate (Figure 4.22). Because it emerged only under strong contextual constraint and disappeared when listeners could not rely on semantic context, the most parsimonious interpretation is that L2 comprehenders engaged in broader context-driven activation during resolution, accompanied by less efficient suppression of semantic competitors once the target had been encountered.

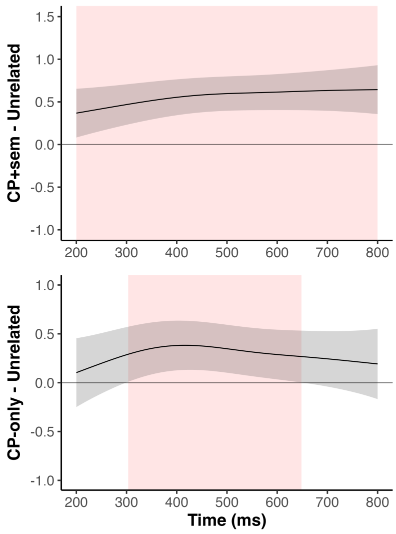
To further explore whether CP-only and CP+Sem competitors showed distinct patterns during word recognition in predictable items, I conducted separate GAMM analyses on the resolution window for each competitor type (see Appendix F.4 for full model summary). Across all participant groups, CP+Sem competitors persisted into the resolution window (L1: 200–800 ms, effect size = 0.553; Dyslexia: 200–800 ms, effect size = 0.563; L2: 200–800 ms, effect size = 0.696; Figure 4.17). This sustained activation likely reflects the fact that CP+Sem items were contextually supported and therefore retained some activation even after the acoustic target onset. In contrast, CP-only competitors showed no significant activation for L1 controls but were activated, albeit more weakly, in both individuals with dyslexia (303–648 ms, effect size = 0.334; Figure 4.17c) and L2 participants (382–800 ms, effect size = 0.432; Figure 4.17b). CP-only items did not show anticipatory activation in any group, indicating that their emergence in the resolution window reflects post-target lexical–semantic spreading rather than prediction. These findings suggest that while CP+Sem competitors naturally carried over into the resolution window for all groups due to their contextual fit, the additional activation of CP-only items in dyslexic and especially L2 comprehenders reflects broader and less selective post-onset lexical activation.



(a) L1 English typical readers



(b) L2 English speakers



(c) Participants with dyslexia

Figure 4.17: Panels show differences in empirical logit-transformed fixation proportions over time between the CP+Sem and unrelated conditions (upper graph in each panel), and CP-only and unrelated conditions (lower graph in each panel) for (a) L1 English typical readers, (b) L2 English speakers, and (c) participants with dyslexia.

4.2.5 Results: Unpredictable Condition Analysis.

To control for general processing differences between groups that might confound the interpretation of predictive processing effects, I analyzed unpredictable sentences where participants heard contexts like "If there is one, click on a picture of..." followed by an unpredictable target word. The critical test lies in the resolution window (post-word onset), which reflects bottom-up lexical access processes that should be equivalent across groups if differences in predictable sentences are truly due to prediction rather than general lexical processing abilities. Following Huettig and Brouwer (2015), who found no differences in lexical access between dyslexic and control participants, I predicted that resolution window effects should not differ between groups in unpredictable sentences. Group differences that emerge only in predictable (but not unpredictable) sentences would, therefore, validate that the main findings reflect prediction-specific processes rather than general lexical access differences.

As expected, model comparison confirmed no systematic competitor effects in the anticipatory window of unpredictable sentences (null vs. complex model: p = 0.626), confirming that there were no systematic group or condition differences during the pre-target period. Analysis of the resolution window of unpredictable contexts (0-1000 ms post-target onset) employed similar model selection procedures. As with predictable sentences, in the resolution time window, the complex model, including both the condition × group interaction and autocorrelation, was strongly favored over simpler alternatives. Comparisons showed a clear improvement over the null model without the interaction (ΔfREML = –12,546.11, Δdf = 33, $p < 0.0001$, AIC difference = 24,800.95), as well as over the model without autocorrelation (ΔfREML = –12,136.33, Δdf = 0, $p < 0.0001$, AIC difference = 24,598.29).

Within-group effects in unpredictable contexts. Analysis of unpredictable contexts provided important insight into bottom-up lexical

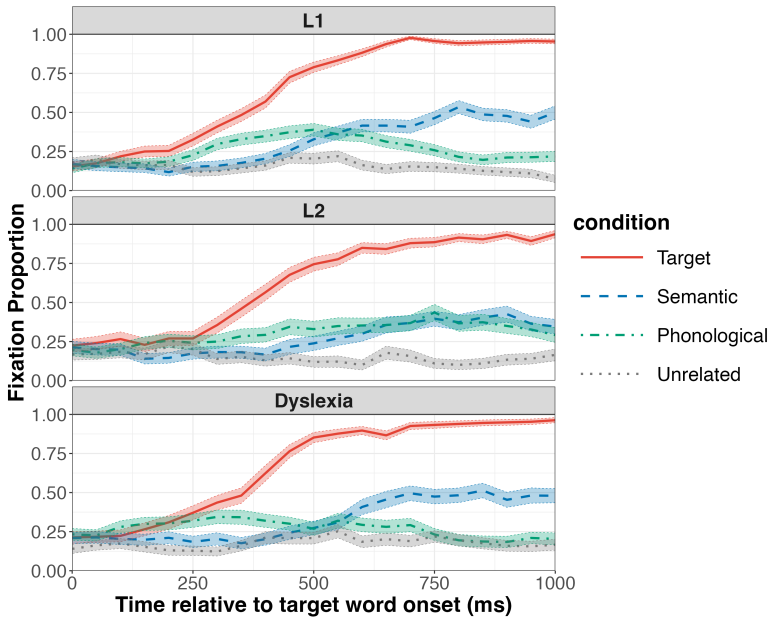


Figure 4.18: Mean fixation proportions on target, semantic competitor, phonological competitor, and unrelated objects in the L1 (top graph), L2 (middle graph), and Dyslexia (bottom graph) groups after the target onset in unpredictable contexts.

access without predictive constraints (see Figure 4.18 for mean fixation proportions plots and Appendix F.5 for the model summary). As expected, in unpredictable contexts, all groups showed successful target recognition with effect patterns emerging only after sufficient acoustic information became available from ∼200 ms onwards (L1: effect size = 1.328; Dyslexia: effect size = 1.280; L2: effect size = 1.359) (see Figures 4.19, 4.20). This contrasts markedly with predictable contexts, where target fixations are already elevated at the start of the resolution window due to anticipatory processing initiated several seconds earlier. Thus, this comparison further confirms that the high target fixations observed at resolution window onset in predictable contexts (Figure 4.15) reflect predictive facilitation rather than purely bottom-up lexical access.

In contrast to predictable items, unpredictable sentences in the resolution window elicited not only semantic competitor effects (L1: effect size = 0.585; Dyslexia: effect size = 0.451; L2: effect size = 0.451), but also phonological competitor activation in all groups (L1: effect size = 0.287; Dyslexia: effect

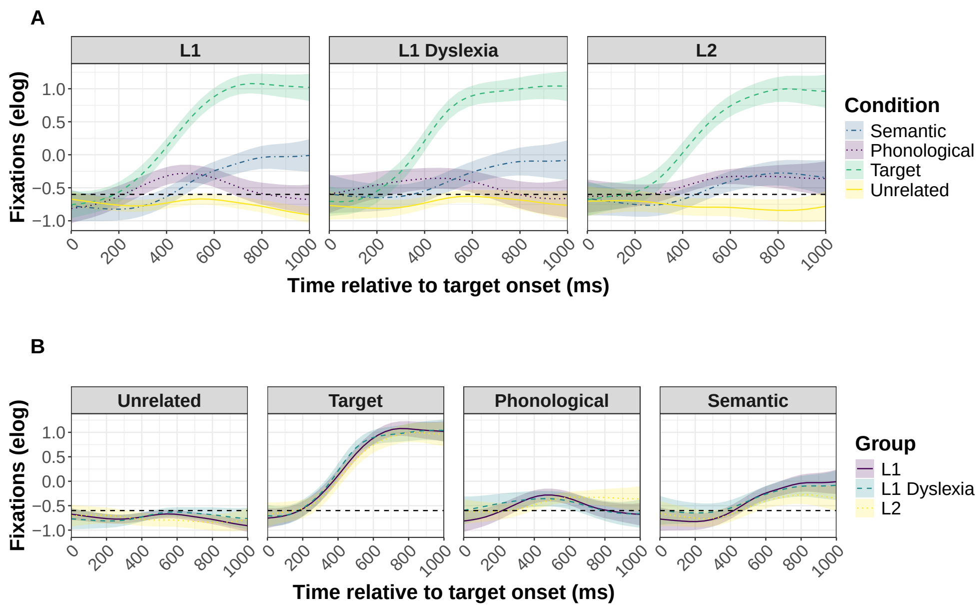
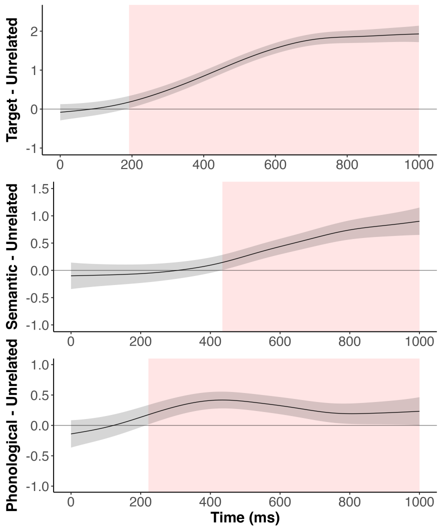
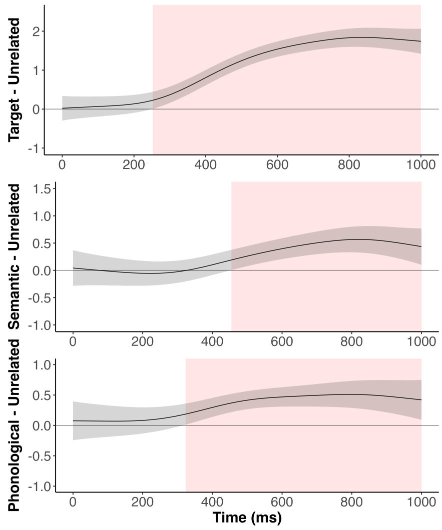


Figure 4.19: GAMM model estimates of looks to the critical object over time in the resolution window of unpredictable items. (A) Comparison by condition for each group. (B) Comparison by group for each condition.

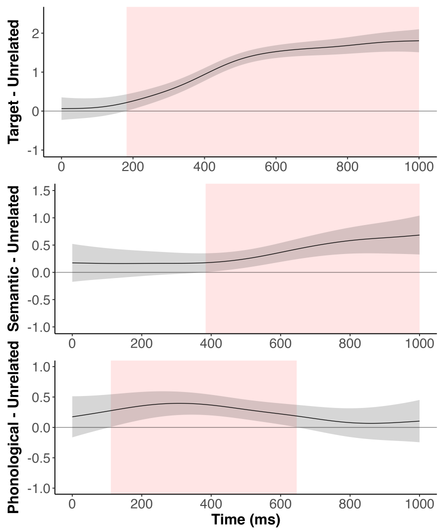
size = 0.322; L2: effect size = 0.431). These phonological effects emerged between ∼200–300 ms and persisted until the end of the time window for L1 and L2 speakers, and until 650 ms for participants with dyslexia. This is an important result, as it supports lexical recognition models that suggest a strong dependence on phonological information during bottom-up processing, particularly when semantic context provides limited constraints (Allopenna et al., 1998; Huettig & McQueen, 2007). Furthermore, this provides extra support to the validity of phonological competitor items used in the study and demonstrates that the absence of their pre-/activation in predictable contexts cannot be explained by inadequate stimulus selection or weak phonological similarity. Rather, the pattern indicates that in semantically constraining environments, top–down predictive processes modulate and effectively suppress the phonological competitor effects that would typically arise during bottom–up lexical recognition.



(a) L1 English typical readers

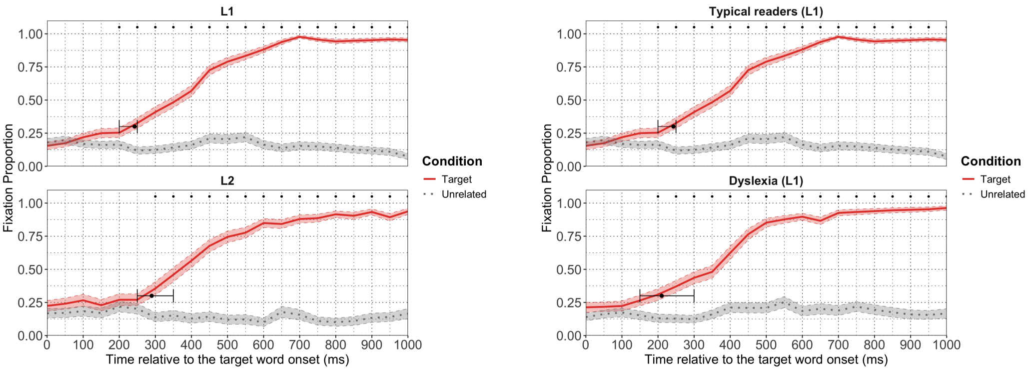


(b) L2 English speakers



(c) Participants with dyslexia

Figure 4.20: Panels show differences in empirical logit-transformed fixation proportions over time between the target and unrelated conditions (upper graph in each panel), semantic and unrelated conditions (middle graph), and phonological and unrelated conditions (lower graph in each panel) for (a) L1 English typical readers, (b) L2 English speakers, and (c) participants with dyslexia in the resolution window of unpredictable items.



(a) (b)

Figure 4.21: Divergence point analysis of mean fixation proportions to the critical word in target vs. unrelated conditions for a) L1 and L2 speakers, and b) typical readers (Control L1) and participants with dyslexia (Dyslexia L1) in unpredictable items.

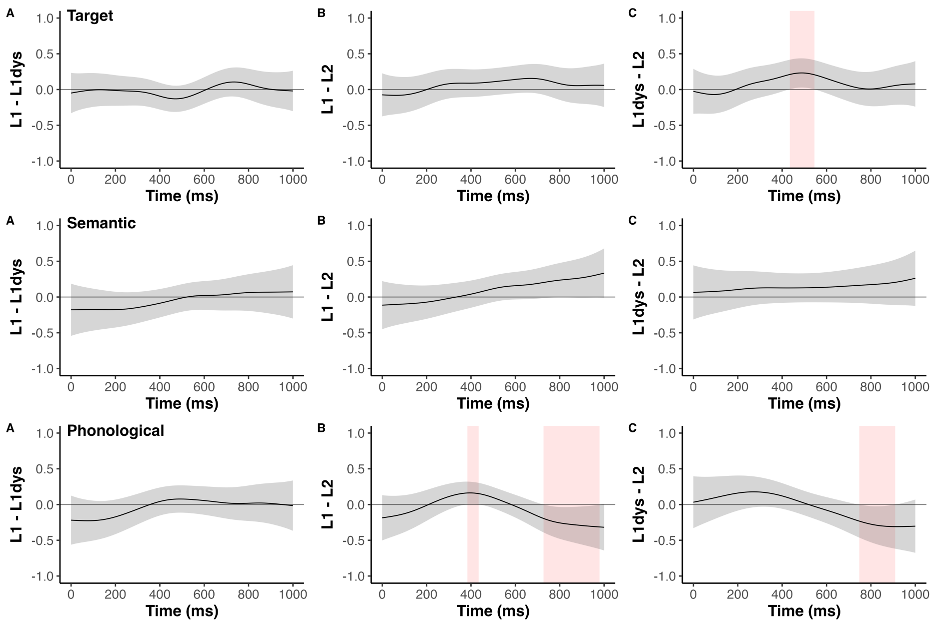
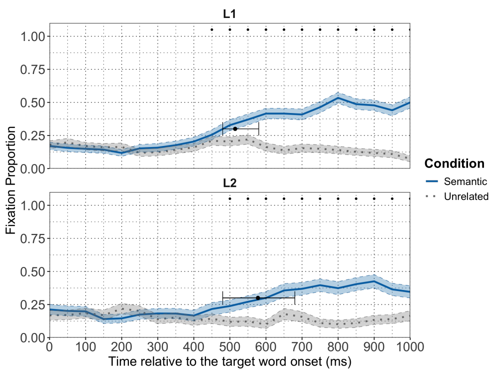
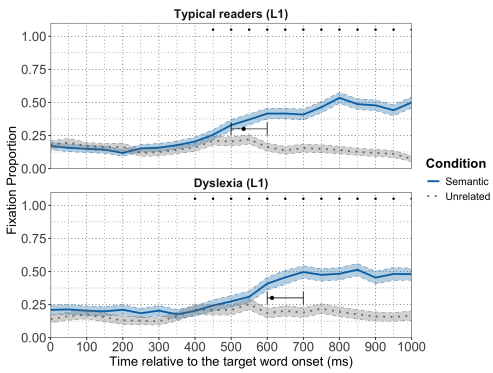


Figure 4.22: Panels show differences in fixation proportions over time between groups in the target (upper panel), semantic (middle panel), and phonological (lower panel) conditions across the resolution time window in unpredictable items for L1 English typical readers vs. participants with dyslexia (leftmost panels), L1 English vs. L2 English participants (middle panels), and dyslexics vs. L2 speakers (rightmost panels).

Between-group comparisons in unpredictable contexts. Both GAMM and divergence point analyses revealed minimal group differences in target recognition when contextual prediction was eliminated. Divergence point analyses showed L1 speakers achieved target recognition at a mean of 243 ms (95% CI: [200, 250]), L2 speakers at 289.4 ms (95% CI: [250, 350]), and the dyslexia group at 210 ms (95% CI: [150, 300]). These timing differences were not statistically significant (all $p > 0.84$; see Figure 4.21), providing evidence that basic word recognition processes remain comparable across groups. GAMM analysis detected a brief period where the dyslexia group showed slightly higher target activation than L2 speakers (447-546 ms, effect size = 0.232; see Figure 4.22). However, this 99 ms window is shorter than the typical 150-200 ms required to launch a saccade in response to auditory input, which could suggest that this difference may reflect random moment-to-moment fluctuations in the data rather than meaningful between-group processing differences.



(a)



(b)

Figure 4.23: DPA of mean fixation proportions to the critical word in semantic vs. unrelated conditions for a) L1 and L2 speakers, and b) typical readers (Control L1) and participants with dyslexia (Dyslexia L1) in unpredictable items.

Semantic competitor activation also showed similar timing across groups in unpredictable contexts. Divergence point analysis revealed L1 speakers initiated semantic effects at a mean of 513 ms (95% CI: [480, 580]), L2 speakers at 577.7 ms (95% CI: [480, 680]), and the dyslexia group at 614.5 ms (95% CI: [600, 700]). None of these timing differences reached statistical significance (all $p > 0.14$; see Figure 4.23), and GAMM analysis detected no significant between-group effects for semantic competitors (Figure 4.22), indicating that semantic competition during bottom-up lexical access follows similar temporal dynamics across all populations. This also indicates that group differences observed in predictable sentences are unlikely to reflect fundamental lexical processing deficits.

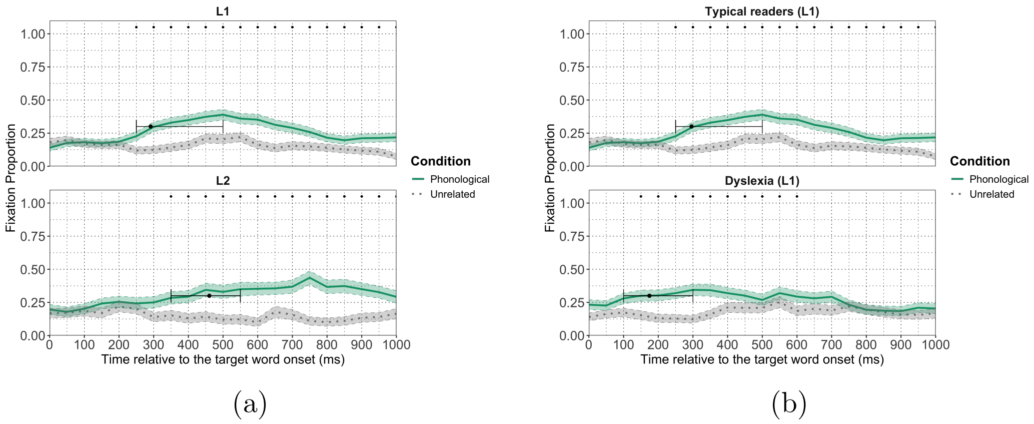


Figure 4.24: DPA of mean fixation proportions to the critical word in phonological vs. unrelated conditions for a) L1 and L2 speakers, and b) typical readers (Control L1) and participants with dyslexia (Dyslexia L1) in unpredictable items.

Phonological competitor divergence points also did not differ reliably between groups. According to DPA, the mean phonological divergence point for L1 speakers was at 293.9 ms (95% CI: [250, 500]), for the dyslexia group at 171.58 ms (95% CI: [100, 300]), and for L2 speakers at 460.6 ms (95% CI: [350, 550]) (see Figure 4.24). Phonological competitor divergence was numerically earlier in dyslexic participants and later in L2 speakers, but these differences did not reach significance (all $p > 0.3$).

While no statistical difference was caught in the divergence of the effects, GAMM analysis revealed significant differences in the duration and resolution of phonological competition (see Figure 4.22). L2 speakers maintained significantly higher phonological competitor activation than both L1 speakers (752-865 ms, effect size = -0.270) and the dyslexia group (723-900 ms, effect size = -0.304) during the later portion of the analysis window. Visual inspection of the difference smooth plots (Figure 4.20) suggests several possible explanations to this pattern: (1) the numerical (though non-significant) delay in L2 phonological onset timing (~170 ms later than other groups) could create a temporal offset that shifts the entire activation profile later in the analysis window; (2) L2 speakers appear to show delayed suppression of phonological competitors: while L1 speakers and individuals with dyslexia demonstrate phonological activation that peaks early and declines as acoustic information disconfirms the phonological cohort, L2 speakers show more sustained activation with a shallower decline slope. This prolonged phonological competition in L2 processing may reflect a qualitatively different organization of the L2 lexicon characterized by form prominence (Gor et al., 2021; Meara, 1978, 1983). The Fuzzy Lexical Representations framework proposes that L2 phonolexical representations are more weakly specified and less sharply differentiated from phonological neighbors (S. Cook et al., 2016; Gor et al., 2021). This reduced specification leads to a lexical competition space that remains active for longer (Gor et al., 2021).

To sum up, the analysis of unpredictable sentences provided validation for the predictive nature of competitor effects observed in predictable contexts. In the anticipatory window, no systematic competitor effects emerged across any group, as expected given the absence of a constraining context. In the resolution window, all groups initiated recognition of the target and its semantic and phonological competitors at comparable times, indicating equivalent onset of lexical access across populations. Importantly, unlike predictable sentences, unpredictable contexts elicited phonological competitor activation in all groups, supporting lexical recognition models that emphasize phonological cohort activation during bottom-up word recognition (Allopenna et al., 1998; Huettig & McQueen, 2007; Luce & Pisoni, 1998; Marslen-Wilson & Welsh, 1978; McClelland & Elman, 1986). This pattern also demonstrates that the absence of phonological prediction in predictable contexts reflects top-down suppression rather than inadequate stimulus design. The results further confirm that group differences in semantic pre-activation observed in predictable sentences stem from prediction efficiency rather than word recognition abilities.

4.3 Discussion

The current experiment provides further evidence of the robustness of contextual prediction across different populations while at the same time reveals some group-specific differences in the temporal dynamics of anticipatory processing during spoken language comprehension. These findings not only replicate but also extend the original results of Ito and Husband (2017) beyond neurotypical L1 populations.

Contextual and semantic competitor prediction. All participant groups demonstrated strong anticipatory fixations to highly predictable target words. The timing of target prediction onset was consistent across groups, with all participants beginning anticipatory fixations approximately 2500-3000 ms before target word onset. This finding confirms that the ability to use sentential context to anticipate upcoming words remains intact even when language processing is challenged by reading disorders or second language status, aligning with theoretical frameworks emphasizing prediction as a core mechanism underlying efficient language processing (Clark, 2013; Dell & Chang, 2014; Pickering & Garrod, 2013). The present results extend established evidence for contextual facilitation in L1 speakers (Altmann & Kamide, 1999; Arai & Keller, 2013; Borovsky et al., 2012; Federmeier, 2007; Ito et al., 2018; Kamide et al., 2003; Van Berkum et al., 2005), and L2 speakers (Chambers & Cooke, 2009; Dijkgraaf et al., 2017; Ito et al., 2018) to adults with dyslexia, for whom evidence remains more tentative, with only one study to date reporting predictive facilitation effects in highly predictable contexts only (Engelhardt et al., 2021).

However, significant temporal differences emerged in the onset of semantic competitor pre-activation. Typical readers demonstrated semantic anticipation the earliest, at around 1900 ms before the target onset. In contrast, dyslexic participants exhibited a significant delay of nearly one second, anticipating at approximately 900 ms before the target onset. L2 participants showed a similar delay of around 1100 ms before the target onset, though this difference was not statistically significant, likely due to high individual variability within this group. This delay, specific to semantic competitor pre-activation (rather than the target word itself), suggests that strong contextual constraints directly facilitate target activation, while the spread of activation to semantically related concepts might function differently and require additional processing time in dyslexic populations, and to some extent in L2 participants. This interpretation fits with the nature of the stimuli: target words were highly predictable continuations of the sentence context, whereas competitors, though semantically related, were not themselves likely continuations. The delay in semantic competitor activation may reflect reduced efficiency in predictive processing, i.e., dyslexic and L2 participants may require additional time and contextual input before activating broader semantic networks.

A further insight came from examining the type of competitors that elicited anticipation. A common limitation of prior studies is that semantic competitor items were typically related not only to the predicted target but also to other content words in the sentence. For instance, in Li et al. (2022), the semantic competitor *eraser* was linked both to the predictable target *schoolbag* and to other items mentioned in the sentence (e.g., *pencil case* and *notebooks* in “After school, I put my pencil case and notebooks into my *schoolbag* and get ready to go home.”). Such overlap makes it unclear whether anticipatory fixations to the competitor reflected activation cascading from the target’s lexical-semantic network or co-activation supported by the broader sentence context. A similar regularity emerged in our materials: competitors that showed anticipatory activation were precisely those linked both to the predicted target and to other contextual elements. This post-hoc observation motivated me to separate competitor items into two subgroups — CP+Sem (competitors related to both high CP target and other contextual words) and CP-only (competitors related only to the target). This analysis revealed that semantic activation occurred exclusively for CP+Sem items. CP-only items did not elicit anticipatory fixations in any group. This finding provides important insights into the constraints governing semantic prediction and suggests that anticipating a highly predictable word does not automatically trigger spreading activation to its lexical-semantic associates. Instead, semantic prediction seems to operate through a constrained mechanism rather than through automatic spreading activation. In other words, predicting *tail* does not necessarily lead to pre-activation of *paw* unless *paw* is also supported by the surrounding context (e.g., dog).

One explanation for the lack of CP-only effects could be that the strong contextual constraint favoring the target word may actively inhibit the activation of its competing lexical candidates, including semantically related words, to reduce processing interference. This relates to constraint-based accounts of language processing reviewed in Chapter 1 (Section 1.1), which assume that multiple representations consistent with the input are activated in parallel, with their relative activation strength determined by probabilistic constraints such as frequency, plausibility, and contextual fit (MacDonald et al., 1994; Trueswell et al., 1994). While such models might predict activation of target-related associates via spreading activation, the present results suggest a more selective weighting process: the target receives maximal activation due to a very strong contextual constraint, CP+Sem competitors achieve intermediate activation based on partial support, and CP-only competitors fail to reach threshold. Thus, pre-activation of a highly predictable word does not automatically lead to the spread of semantic activation.

Exploratory correlational analyses revealed group-specific associations between individual differences and CP+Sem anticipation timing. In the dyslexia group, higher receptive vocabulary was associated with earlier semantic competitor anticipation, consistent with evidence that vocabulary knowledge is a key predictor of anticipatory language processing in adults (Hintz et al., 2017; Huettig & Pickering, 2019; Rommers et al., 2015). In the L2 group, greater print exposure was associated with earlier CP+Sem anticipation, suggesting that accumulated reading experience may support the development of lexical-semantic networks that enable more efficient context-driven activation in this population (cf. Favier et al., 2021; Mishra et al., 2012, for evidence in L1 populations). No associations were observed in L1 typical readers, possibly reflecting the more restricted range of both predictor and outcome variables in this group, or the greater automaticity of predictive processing in typical native speakers. Together, these findings suggest that CP+Sem anticipation timing is not consistently explained by any single individual-difference measure across participants. Rather, lexical experience appears to influence prediction efficiency most visibly when processing is less entrenched, with different indices — vocabulary breadth versus print exposure — being most relevant depending on the population. Given their exploratory nature and modest effect sizes, these results should be interpreted with caution.

Phonological form prediction. Turning to phonological prediction, the current study found no evidence that the phonological form of predicted words was pre-activated, even under conditions optimized for detecting such effects (faster speech rate, high predictability). This replicates Experiment 1a and extends the null findings of Ito and Husband (2017), contrasting earlier reports of phonological pre-activation (DeLong et al., 2005; Ito et al., 2018, 2020; Li & Qu, 2023; Li et al., 2022). Moreover, the absence of phonological anticipation should not have been due to stimulus design: in unpredictable contexts, all groups displayed strong phonological competition, confirming the validity of phonological competitors. Converging evidence from a production study by Drake and Corley (2014) points in the same direction: when a sentence strongly constrains a specific word, naming is facilitated only when the picture name exactly matches the predicted word, with no onset/rime benefits for phonological competitors.

One important caveat in interpreting the absence of phonological competitor activation is that it does not necessarily imply that phonological form was not pre-activated. An alternative and somewhat opposing interpretation is that the target's phonological form was so strongly anticipated that competitors were actively suppressed, preventing any measurable activation. This possibility mirrors the CP-only vs. CP+Sem distinction observed in the semantic domain: while CP-only competitors, those sharing semantic features only with the target word itself, failed to elicit activation in highly predictive contexts, CP+Sem competitors did achieve measurable activation, albeit weaker than the target. This contrast suggests that the predictive system does not suppress all non-target candidates indiscriminately, but rather selectively suppresses those lacking contextual support, while permitting residual activation for items that are also related to other contextual words in the sentence. Similarly, recent work by Haeuser and Borovsky (2024) shows that strongly predicted targets suppress phonological cohorts during word recognition. This is consistent with Clark's (2013) predictive coding theory, which assumes that the brain continuously generates predictions about upcoming input and updates them by minimizing prediction error. From this perspective, the absence of phonological competitor effects in the present data may reflect not a lack of form prediction but the efficiency of prediction-induced suppression mechanisms: the system only pre-activates what is useful to reduce uncertainty. Indeed, interactive accounts such as TRACE explicitly allow top–down contextual information to influence and modulate cohort activation in real time (see Section 1.1). In our case, if the target is already near-certain (e.g., 95–100% cloze probability), then pre-activating a wide set of semantic or phonological associates wouldn't reduce prediction error but would consume processing resources. Thus, the system "economizes" by focusing prediction narrowly on the target. Only when constraints are weak (as in unpredictable contexts) does the system rely on bottom–up mechanisms that activate phonological cohorts (Allopenna et al., 1998; Huettig & McQueen, 2007). In this way, the absence of phonological prediction can be interpreted not as a limitation of the predictive system, but as an outcome of its efficiency. The system economizes resources by deploying spreading activation of the phonological cohort only when it is needed to support recognition, rather than when context nearly guarantees the target. This selective deployment is consistent with Hawkins' (2004, 2014) broader argument that the language processing system maximizes online efficiency by limiting activation to representations that serve immediate comprehension needs, since preactivating unnecessary candidates would consume resources without reducing uncertainty.

Unpredictable items. The results from unpredictable items serve as an important point of comparison. Across all groups, when the target word was not predictable from the preceding context, participants displayed strong lexical competition effects, with both phonological and semantic competitors becoming activated after target onset. This pattern replicates classic findings from visual-world studies (Allopenna et al., 1998; Huettig & McQueen, 2007), where lexical access proceeds through bottom–up input, leading to the initial activation of phonological cohorts and, shortly after, semantically related words. These competition effects were similar across L1, dyslexic, and L2 participants, suggesting that the basic architecture of lexical activation and competition is intact across groups. This semantic activation during bottom-up lexical access aligns with previous findings showing that lexical processing automatically spreads to conceptually related items during word recognition (Huettig & Altmann, 2005; Yee & Sedivy, 2006).

This stands in contrast to the results for highly predictable items, where anticipatory prediction of the target significantly reduced competitor activation. The system seems to function in two different modes. When there are no strong contextual constraints, lexical access is driven from the bottom up, and multiple competitors are activated simultaneously. However, when strong contextual constraints are present, lexical access operates from the top down, suppressing competitors in favor of the predicted target.

Resolution window dynamics in predictable items. Analyzing the resolution window in predictable items revealed interesting insights into how competitor activation occurs once bottom-up input is available during the comprehension of highly predictable sentences. For L1 controls, CP+Sem competitors remained active into the resolution window, whereas CP-only competitors showed no activation, which looks as though only contextually licensed alternatives survive integration. By contrast, both individuals with dyslexia and L2 speakers showed not only CP+Sem activation but also statistically reliable activation of CP-only items. This CP-only effect emerged later (around 300 ms after the target onset) and was weaker than the CP+Sem activation, which started well before the target, indicating that the CP-only effect did not reflect anticipatory prediction but rather post-onset spreading activation from the encountered target. This highlights a key difference across populations: while L1 controls selectively maintain activation only for competitors supported by contextual cues, dyslexic and L2 listeners activate a broader set of semantic associates once the target is processed. This may reflect greater lexical uncertainty during spoken language processing in these groups.

The persistence of CP+Sem competitors into the resolution window, contrasted with the absence of CP-only activation in L1 controls, can be understood in terms of the interplay between local semantic priming and global sentential prediction. In the current study, CP+Sem items were linked both to the predictable target and to other lexical items in the preceding context. Although they were implausible continuations compared to the target, this convergence of multiple local semantic relations endowed them with sufficient activation strength to survive into the integration phase (though with a much smaller effect size compared to highly predictable targets). By contrast, CP-only items were linked exclusively to the predicted target and lacked independent contextual support; as a result, they never crossed the threshold for anticipatory or integration-phase activation in L1 listeners. This account resonates with the constraint-based frameworks reviewed in Chapter 1 (Section 1.1), which posit that multiple probabilistic cues jointly determine activation strength during comprehension (MacDonald et al., 1994; Trueswell et al., 1994). Empirical work by Kukona et al. (2011) illustrates this graded interplay of predictive and priming mechanisms. In their experiments, both sentence-level predictability and verb-based thematic priming influenced anticipatory looks, with strong competitor activation when supported by local semantic associations, even if those competitors were not the most probable continuations. Similarly, Hoeks et al. (2004) found that strong sentence-level constraint did not fully override local lexico-semantic fit, as incongruent items still elicited N400 effects in constraining sentences. Taken together, these studies suggest that local semantic relations can sustain activation of semantically related competitors despite being incongruent with global contextual predictability. This is consistent with the persistence of CP+Sem activation during the resolution window in the current study.

The absence of CP-only activation in the population of L1 typical readers shows that semantic prediction is not an automatic by-product of target anticipation but depends on converging contextual support. Dyslexic and L2 participants, however, displayed additional CP-only activation post-onset, which I interpret as lexical-semantic spreading from the encountered target (similar to the semantic activation in the unpredictable contexts) rather than anticipatory prediction. This broader post-onset activation likely reflects noisier or less selective inhibition and greater lexical uncertainty in these populations. From the perspective of Kukona et al.'s (2011) distinction between "exploitation" and "exploration", L1 typical readers can afford to maximize by committing to the highly predictable target and suppressing unsupported alternatives, because their lexical representations are stable and processing resources are sufficient to handle unexpected input if it arises. By contrast, dyslexic and L2 participants may adopt a more exploratory strategy, relying more on the bottom-up input and maintaining access to additional competitors even after target onset. This exploration ensures flexibility under conditions of processing difficulty or limited linguistic experience.

Summary. In summary, the present study confirms that prediction of high-cloze targets is a robust phenomenon across groups, but it also provides new insights into the extent and selectivity of prediction. First, the CP-only vs. CP+Sem contrast demonstrates that anticipating a predictable word does not automatically spread activation to its semantic associates. Instead, semantic prediction emerges only when competitors are supported by converging contextual cues. Second, all groups showed semantic competitor activation, but typical L1 speakers engaged this process earlier. Adults with dyslexia were significantly delayed, while L2 participants showed a numerically later onset with high variability, such that the difference did not reach statistical significance. This delay in individuals with dyslexia suggests reduced efficiency in predictive processing, where additional time and contextual input are required before activating broader semantic networks. Exploratory correlational analyses further suggested that within-group variability in CP+Sem anticipation timing was partially associated with vocabulary knowledge in the dyslexia group and print exposure in the L2 group. Third, the absence of phonological pre-activation could be explained by two competing theories: 1) the phonological form of predicted targets does not get activated, or 2) phonological activation does happen, but similar to the CP-only finding, the system suppresses unnecessary activation of phonological cohorts in highly constraining contexts, conserving resources and maximizing efficiency. Finally, resolution-window analyses revealed that individuals with dyslexia and L2 speakers, but not L1 controls, additionally activated CP-only competitors post-onset, reflecting broader and less selective lexical-semantic spreading once the target is processed.

Group differences can be understood within the framework of exploitation versus exploration in predictive processing (Kukona et al., 2011). L1 listeners adopt an exploitation strategy, committing to the highly predictable target and suppressing unsupported alternatives, reflecting efficient maximization under strong contextual constraint. By contrast, dyslexic and L2 participants display a more exploratory profile, maintaining activation of additional competitors post-onset. This broader activation may reflect a combination of noisier or less selective inhibition mechanisms and greater lexical uncertainty, which together reduce efficiency in narrowing the competitor set. However, such exploration may also represent an adaptive strategy under conditions of processing difficulty or limited linguistic experience, providing flexibility when the input is less predictable. Future research could test this account directly by manipulating contextual constraint, for instance, by examining whether lower-cloze sentences increase exploratory activation of both semantic and phonological competitors across groups.

Chapter 5

General Discussion and Conclusions

5.1 Methodological Considerations for Cross-Experiment Comparison

Although Experiments 1a and 2 address similar theoretical questions about prediction in spoken language comprehension, design differences preclude direct statistical comparisons. The two experiments were not initially intended to be direct replications of one another. Rather, they were designed to examine predictive processing from different angles and in different linguistic environments. Comparing them, therefore, provides complementary perspectives. The goal of this section is to specify what is comparable, what is not, and how I interpret cross-study patterns in light of those constraints. Both studies used the Visual World Paradigm with eye-tracking, recruited parallel participant groups (L1 typical readers, L1 readers with dyslexia, L2 speakers), and employed highly predictable sentences (Italian: mean cloze = 92.4%; English: mean cloze = 95%). Phonological cohort competitors had a comparable proportion of overlap across languages (Italian = 0.40; English = 0.49). In both experiments, the language of instructions and testing corresponded to the native language of the L1 groups (Italian in Experiment 1, English in Experiment 2). There were, however, several methodological differences:

1. **Target position and sentence structure.** Experiment 1a employed uniform sentence structures with targets always in sentence-final position. This design likely created relatively predictable timing windows for anticipatory processing, potentially facilitating stronger or earlier prediction effects. Prior work has shown that listeners can adapt their predictive strategies based on the statistical regularities of an experimental context (e.g., Brothers et al., 2017; Fine et al., 2013). Participants in Experiment 1a may therefore have developed expectations not only about content but also about the timing of critical items. In contrast, Experiment 2 (following Ito & Husband, 2017) implemented more variable syntactic structures, with targets appearing in different sentence positions. This variability may have required more naturalistic, adaptive processing, allowing for more flexible prediction over time. While the more consistent structure of Experiment 1a likely promoted stronger or earlier anticipatory effects, this does not imply a qualitative difference in the underlying cognitive mechanisms. Prediction is widely viewed as a general, resource-dependent mechanism that flexibly adapts to contextual demands without changing its core nature (Huettig, 2015; Pickering & Gambi, 2018).

2. **Task demands.** In Experiment 1a, participants were instructed to look at the named object, while in Experiment 2, they additionally clicked on it (or on the background if absent). The click requirement introduced a light motor-decision component that could, in principle, modulate post-onset fixation patterns. However, research using eye-tracking in sequence learning tasks has demonstrated that anticipatory eye movements occur with similar frequency and patterns regardless of whether participants make concurrent manual responses (Marcus et al., 2006). Thus, the click requirement should not affect anticipatory fixations.

3. **Speech rate.** Stimuli also differed in average speech rate (Italian = 5.0 syll/s; English = 3.5 syll/s), reflecting the natural prosodic characteristics of the two languages. Because speech unfolded at different rates, absolute fixation latencies cannot be directly comparable.
4. **Preview time.** In Experiment 1a, pictures appeared 2000 ms before the target word onset, whereas in Experiment 2 they appeared 2000 ms before sentence onset, resulting in a longer preview period overall. While a longer preview could influence the strength of phonological competition, both designs provided ample time for visual encoding of the display (Huettig & McQueen, 2007) and preserved the predictive contexts necessary for observing anticipatory eye movements.
5. **Apparatus and sampling.** Different experimental platforms and eye-tracking systems were used. Experiment 1a was designed in Open Sesame, an open-source tool based on Python (Mathôt et al., 2012), with eye movements recorded using a Gazepoint eyetracker sampling at 60 Hz, while Experiment 2 was built in SR Research Experiment builder and used an EyeLink 1000 Desktop mount eye-tracker for gaze recording sampling at 500 Hz. These systems differ in sampling resolution; therefore, as noted above, all analyses were conducted separately within each experiment, and results are compared only at the interpretive level (i.e., overall patterns rather than absolute values).
6. **Participant characteristics.** Beyond methodological differences in experimental design, the two studies also differed in the sociolinguistic contexts that shaped participants' language experience. The Tyrolean L1 speakers in Experiment 1 were tested in their L2 (Italian) in South Tyrol, a territory where both languages can coexist in daily life (see Section 1.5). These participants maintained regular use of both languages, with Tyrolean remaining their dominant language in terms of daily exposure (67% vs. 20% for Italian; see Table 2.2). In contrast, the L2 participants in Experiment 2 were speakers of Romance languages living in an English-dominant environment (the UK). These

participants showed reversed dominance patterns, with higher daily L2 English exposure (72%) than L1 exposure (26%; see Table 4.3).

Importantly, these methodological differences do not compromise the comparability of core prediction effects, as both experiments maintained the essential design for investigating anticipatory processing with high cloze probability contexts. On the contrary, finding that prediction patterns emerge across different structural implementations would strengthen the overall conclusions, suggesting robustness of the observed effects. In light of these considerations, cross-experiment comparisons are interpreted at the level of general patterns by "reading the two studies side-by-side". The focus is on whether the same general patterns replicate: strong anticipation of high-cloze targets, absence of anticipatory phonological competitor effects, and delayed anticipatory effects in L2 and dyslexia groups. By contrast, I do not compare the absolute size of effects or their precise millisecond latencies, since speech rate, task demands, and hardware differed. Instead, I concentrate on whether comparable effects are present or absent within predictive and resolution time windows. In the next section, I use these principles to bring together results across linguistic levels and participant groups.

5.2 Integrated Findings Across Experiments

5.2.1 Contextual (Target) Prediction Across Experiments and Groups

Across both experiments, the results confirm the already established finding that listeners use contextual cues to anticipate upcoming referents before their acoustic onset. This is well documented in L1 processing (e.g., Altmann and Kamide, 1999; Borovsky et al., 2012; Kamide et al., 2003; Kutas and Federmeier, 2011), and has also been shown to extend to L2 comprehension (and, more rarely, to populations with dyslexia) under appropriate conditions (see Section 1.4; e.g., Borovsky et al., 2012; Chambers and Cooke, 2009;

Dijkgraaf et al., 2017; Ito et al., 2018; Tagliani et al., 2025; Van Bergen and Flecken, 2017 for L2, and Engelhardt et al., 2021 for dyslexia).

In line with this broader literature, both Experiment 1a and Experiment 2 showed anticipatory looks to predictable targets across all groups, indicating that L2 comprehenders and readers with dyslexia engaged predictive mechanisms rather than adopting a purely bottom-up strategy. Divergence onsets were broadly comparable across L1, L2, and dyslexia groups, suggesting that target anticipation can unfold on a similar timeline irrespective of language status or reading profile. At the same time, the two experiments differed in the magnitude of anticipatory fixations to targets across L1 and L2 speakers. In Experiment 1a, Tyrolean participants listening in their L2 Italian displayed a shallower build-up of looks to the predictable target compared to L1 Italian speakers, with lower fixation proportions from around 800 ms before target onset and a later divergence point (about 250 ms later than L1 speakers). This timing difference reached the conventional threshold for statistical significance ($p = .05$), but should be interpreted with caution given overlapping confidence intervals and variability in the estimates. By contrast, in Experiment 2, L2 English speakers patterned more closely with the L1 group, showing both comparable divergence and similar overall proportions of target fixations in the anticipatory window. Adults with dyslexia displayed eye movement patterns broadly similar to those of typical readers across both experiments: they showed reliable anticipatory looks and no consistent reduction in fixation proportions, indicating no pronounced difficulties with prediction for high-cloze contextual constraints (consistent with Engelhardt et al., 2021).

So, why did L2 speakers show weaker anticipatory looks in the Italian experiment but not in the English one? At least two factors likely contribute, which are not mutually exclusive. *(1) Contextual constraint.* The Italian materials had slightly lower cloze probability (mean 92.4%) than the English materials (mean 95%). Although both values fall within the high-constraint range and this difference alone is unlikely to fully account for the observed pattern, it is possible that even a modest reduction in constraint affected the L2 group, whose anticipatory looks may reflect greater sensitivity to contextual uncertainty than those of L1 speakers. *(2) Bilingual dominance and proficiency profiles.* As mentioned above (Section 5.1), the two L2 groups differed in language usage patterns. Tyrolean–Italian bilinguals reported higher daily Tyrolean L1 use and lower Italian L2 exposure and proficiency (Tyrolean: ∼67% daily L1, ∼20% daily Italian; L2 Italian self-rated mean proficiency = 4.3 vs. L1 Tyrolean mean = 5.6), whereas the English L2 group reported higher daily L2 use and nearly balanced proficiencies (English L2 daily exposure ∼72%, L1 ∼26%; L2 English self-rating mean proficiency = 5.1, L1 mean = 5.5). These dominance patterns might explain why the Italian L2 group displayed weaker anticipatory proportions. With less exposure and lower proficiency, their lexical–semantic representations in Italian may be less strongly entrenched, leading to more diffuse activation, higher uncertainty, and weaker target commitment. This interpretation is consistent with evidence that (a) vocabulary knowledge supports stronger prediction, (b) proficient bilinguals can match L1 prediction under strong constraints, and (c) linguistic experience and higher proficiency boost L2 prediction (Borovsky et al., 2012; Dijkgraaf et al., 2017; Tagliani et al., 2025; Van Bergen & Flecken, 2017).

Taken together, the findings from both experiments suggest that contextual prediction is a robust mechanism observed across languages and participant groups, extending to both L2 speakers and individuals with dyslexia. Under strong contextual constraint, such anticipatory activation emerges reliably even when proficiency or processing efficiency is reduced. However, its strength varies: L2 listeners anticipate upcoming referents with broadly native-like timing, but the certainty of their predictions scales with contextual constraint, dominance patterns, and proficiency.

5.2.2 Semantic Competitor Activation

Semantic competitor pre-activation was examined only in Experiment 2, but the results provide important insights into how prediction unfolds across groups. While semantic pre-activation has been demonstrated in a handful of L1 studies (e.g., reduced N400s to within-category violations in reading: Federmeier and Kutas, 1999; Ito et al., 2016; anticipatory fixations to semantic competitors in visual-world paradigms: Ito and Husband, 2017; Li and Qu, 2023; Li et al., 2022; facilitation of related but unpredicted words in reading: Frisson et al., 2017), direct evidence for such effects in L2 comprehension is extremely limited (see Dijkgraaf et al., 2019). Against this background, the present study adds novel evidence by showing that participants across all groups directed anticipatory looks not only to highly predictable targets but also to semantically related competitors. This supports the view that prediction is not restricted to a single lexical item but can, under certain conditions, extend to related concepts.

At the same time, significant group differences emerged in the timing of semantic competitor pre-activation. Typical readers anticipated competitors earliest (about 1900 ms before target onset), whereas both L2 participants and readers with dyslexia showed delays of roughly one second. For L2 participants, this effect did not reach statistical significance, likely due to variability within the group, but participants with dyslexia consistently diverged from unrelated items nearly a second later than controls. These delays are notable because they were specific to semantic competitor activation: all groups predicted the highly constraining target words with similar timing, but neurotypical native speakers were the earliest to extend this prediction to semantically related alternatives.

For L2 participants, the difference in competitor activation did not reach statistical significance, so it should be interpreted with caution. Nevertheless, the tendency toward a later onset of semantic competitor anticipation is consistent with prior work suggesting that semantic prediction in a second language may be slightly delayed relative to the L1. Dijkgraaf et al. (2019) observed weaker and slower semantic competitor effects in Dutch–English bilinguals, and interpreted this in line with the temporal delay assumption of the BIA+ model of bilingual word recognition (Dijkstra & Van Heuven, 2002). According to this account, lower subjective frequency of L2 words leads to slower activation of their forms and, as a consequence, to a delayed spread of activation to semantics. Dijkgraaf et al. (2019) also emphasize that there is no qualitative difference between semantic prediction in the L1 and L2: both activate the same underlying mechanisms but with subtle quantitative differences. The present findings are compatible with this view. Our English L2 participants were highly proficient and reported significant daily exposure, which may explain why they demonstrated similar target prediction to L1 speakers. However, semantic competitor pre-activation might place additional processing demands, and even small differences in lexico-semantic mappings (Gollan et al., 2008; Gollan et al., 2005) and word form activation may have led to a narrower or slightly slower spread of activation in the L2.

For dyslexia, the present results represent the first direct evidence concerning semantic competitor pre-activation. Previous studies on prediction in dyslexia have focused on other domains. Engelhardt et al. (2021) used a response-time paradigm to test whether highly predictable targets were pre-activated, and found that adults with dyslexia did anticipate targets in high-cloze sentences, but their predictions were less reliable in lower-cloze contexts. Huettig and Brouwer (2015), by contrast, examined morphosyntactic prediction via grammatical gender cues and reported delayed anticipatory effects in adults with dyslexia. To date, however, no study has assessed whether individuals with dyslexia pre-activate semantic competitors of an expected word. The current findings, therefore, extend the scope of research on predictive processing in dyslexia by showing that, although adults with dyslexia did anticipate highly predictable targets on time, they were significantly delayed in anticipating semantically related competitors. This suggests that while strong contextual support is sufficient to elicit target prediction, the spread of activation to broader semantic networks is less efficient in dyslexia.

A further insight came from examining the type of competitors that elicited anticipation. Dijkgraaf et al. (2019) noted a limitation in their study, noting that semantic competitor effects could have arisen either through spreading activation from the pre-activated target or through similarity with other semantically related words in the sentence. This distinction aligns closely with the present findings. Only CP+Sem competitors (related both to the highly predictable target and to other contextual elements) elicited anticipatory looks. CP-only competitors, linked solely to the target, did not attract anticipatory fixations in any group. This suggests that strong contextual constraint do not automatically trigger activation of all lexical associates of the predicted word. Instead, competitor activation appears to be selective, emerging when multiple cues are present. This pattern aligns with constraint-based accounts of sentence processing (MacDonald et al., 1994; Trueswell et al., 1994), in which candidate activations are weighted by their degree of contextual support: the target receives maximal activation, CP+Sem competitors achieve intermediate activation, and CP-only competitors fail to reach threshold.

Thus, the findings suggest that predicting a single highly probable target is relatively robust across populations, whereas extending prediction to semantically related alternatives is more demanding and therefore more vulnerable to group-level differences. The fact that all groups showed the same qualitative pattern of competitor activation — facilitating CP+Sem but not CP-only competitors — suggests that the predictive mechanism itself operates similarly across groups, with differences arising in efficiency or speed rather than in the fundamental architecture (cf. Dijkgraaf et al., 2019; Engelhardt et al., 2021; Kaan, 2014). As shown in Section 4.2.3, the timing of semantic competitor activation was not predicted by any cognitive or reading-related individual-differences measure in the present battery for the sample as a whole. However, group-specific analyses revealed the following pattern: within the dyslexia group, higher receptive vocabulary was associated with earlier CP+Sem anticipation, while within the L2 group, greater print exposure predicted earlier activation. The absence of significant associations in typical L1 readers may reflect the more restricted variance and greater automaticity of predictive processing in this group. Taken together, these findings suggest that lexical experience modulates prediction efficiency most visibly in contexts where that experience is less uniformly high across individuals (L2 speakers and individuals with dyslexia). This pattern is consistent with evidence that vocabulary knowledge and cumulative language exposure support anticipatory language processing more broadly (Favier et al., 2021; Hintz et al., 2017; Huettig & Pickering, 2019; Rommers et al., 2015).

Moreover, the finding that only CP+Sem competitors elicited anticipatory looks indicates that semantic spread is not automatic: unsupported competitors appear to be suppressed in high-cloze contexts. This suggests that predictive processing is selective, weighting candidates by contextual support rather than indiscriminately pre-activating all associates of the predicted word. The analysis of the resolution window further highlights these dynamics. Once bottom-up input became available, L1 controls maintained activation only for CP+Sem competitors, whereas CP-only competitors remained at baseline. Although the CP+Sem items were not viable continuations, they had local support, linking to both the predicted target and other words in the previous context. This support seems sufficient to sustain their activation briefly into integration. This selectivity for locally supported competitors in L1 aligns with accounts in which local semantic fit can maintain activation despite low global plausibility (Federmeier et al., 2007; Ito et al., 2018). By contrast, both L2 participants and adults with dyslexia showed not only CP+Sem activation but also reliable post-onset activation of CP-only competitors. This effect emerged later (roughly 300 ms after target onset) and was weaker than the CP+Sem activation, consistent with lexical–semantic spreading from the encountered target during routine word recognition rather than anticipatory prediction. From a theoretical standpoint, these results support the view that L1 listeners use strong contextual cues to maximize efficiency by committing to the predicted target, whereas L2 and dyslexic listeners engage in a more exploratory strategy, tolerating additional competitor activation as a safeguard against processing uncertainty (Kukona et al., 2011).

5.2.3 Phonological Competitor Activation

Across both studies, I found no evidence that the phonological form of highly predictable words was pre-activated before acoustic onset. In Experiment 1a, anticipatory looks to phonological competitors did not exceed unrelated baselines in any group, and phonological competitors remained inactive during word recognition in the resolution window of predictable items. A parallel pattern emerged in Experiment 2: anticipatory phonological effects were absent, and phonological competitors did not show above-baseline activation during the resolution window of predictable sentences. Although some studies report phonological pre-activation, a recent meta-analysis by Ito (2024) indicates the effects are small and highly design-sensitive, modulated by contextual constraint, preview timing, and degree of overlap, with both null and positive anticipatory findings reported (nulls: Ito and Husband, 2017; Nieuwland et al., 2018; null in L2 population: Ito et al., 2017b; positives: DeLong et al., 2005; Ito et al., 2018, 2020; Li and Qu, 2023; Li et al., 2022).

A key limitation of Experiment 1a is the absence of an unpredictable control to test for general cohort activation during word recognition. Thus, null effects in predictable sentences could reflect insufficient form overlap between targets and competitors, as overlap magnitude is a known moderator of cohort activation (Ito, 2024). Experiment 2 addressed this by including a block with unpredictable sentences ("If there is one, click on a picture of a ..."). In these contexts, all groups showed phonological cohort competition after target onset, confirming that the phonological competitor manipulation and stimuli were effective and that null findings in predictable sentences are not attributable to invalid competitors. Importantly, in both experiments, phonological overlap and temporal alignment were defined with respect to the acoustic signal actually perceived by listeners, based on broad phonetic transcriptions of the recorded utterances (Heselwood, 2013) rather than dictionary forms — an aspect that is rarely reported in the prediction literature. Explicit control over these parameters is important for interpreting phonological prediction effects, given growing evidence that speech rate modulates predictive processing, particularly in L2 speakers and populations with reduced processing efficiency (Fernandez, Hadley, et al., 2025; Fernandez et al., 2020).

Simulation-based power analyses indicated that Experiment 1a had low power (estimated power = 22.8%), primarily due to high residual variance and a limited number of effective observations per condition, whereas Experiment 2 provided a considerably more sensitive test (estimated power = 86.9%). This represents a substantial improvement over Experiment 1a, reflecting both the larger sample and lower residual variance afforded by the higher temporal resolution of the EyeLink 1000. Moreover, the convergence of null phonological prediction effects across two independent experiments, each with different samples and designs, adds evidential weight to this pattern beyond what either study could provide alone.

I consider two possible explanations for null phonological effects in predictable contexts. *(1) Lemma-only prediction.* On this view, prediction remains at the semantic/lemma level under strong constraint, so no cohort-level phonological effects can arise (Nieuwland et al., 2018). This remains a theoretical possibility for comprehension, but evidence from production studies argue against a lemma-only view: in a sentence-to-picture naming task, Drake and Corley (2014) showed that strongly constraining contexts sped naming when the picture's name exactly matched the predicted word, indicating that predicted words reach the phonological level (see pages 45-46). *(2) Predictive suppression under high constraint.* Alternatively, the target's phonological form is pre-activated, but strong context quickly narrows processing to the expected word and down-weights activation of phonological neighbors, leaving no measurable cohort advantage even after word onset (see also Haeuser & Borovsky, 2024). Experiment 2 supports the suppression-compatible view since identical cohort competitors were activated in unpredictable sentences for all groups but not in highly predictable contexts. Evidence from production studies complements this pattern: sentence constraint facilitated naming only when the pictured object's name exactly matches the predicted word, with no onset/rhyme benefits for phonological neighbors (Drake & Corley, 2014). Distinguishing between these two accounts will likely require other methods like reaction-time paradigms that parallel production evidence without requiring articulation (see Future Directions 5.3.2). Both accounts align with interactive feedback models, in which top-down context influences and narrows lexical competition (e.g., TRACE), as well as with predictive-coding theories that interpret this narrowing as a reduction in prediction error (Clark, 2013; McClelland & Rumelhart, 1981).

In short, the present results suggest that predictive processing during spoken comprehension rarely extends to pre-activating phonological competitors of highly predictable targets under strong contextual constraint. When constraint is high, the system narrows activation to the target and down-weights competing cohorts, producing no detectable phonological competitor looks even post-onset. This pattern aligns with the CP+Sem/CP-only findings in the semantic domain: only competitors with additional local support (CP+Sem) get preactivated, whereas unsupported alternatives (CP-only) do not. When constraint is low (unpredictable contexts), bottom-up dynamics re-emerge uniformly across groups activating both phonological cohorts, as well as semantic competitors. Together, these findings argue for a selective use of phonological activation: it is readily observed as part of bottom-up access, but does not routinely arise as a predictive mechanism in highly constraining sentences. This pattern aligns with Hawkins' (2004, 2014) proposal that efficient language processing systems maximize online access to what is needed while minimizing activation of what is not.

5.2.4 Morphosyntactic Prediction

The results of Experiment 1b expand our understanding of predictive processing by targeting morphosyntactic cues, specifically grammatical gender. This type of prediction was tested only in the Italian experiment, as English lacks a system of nominal gender agreement on determiners necessary for an equivalent manipulation. Unlike contextual and semantic prediction, which reliably generated anticipatory eye movements, gender-marked determiners in Italian did not elicit anticipatory looks across any group. The absence of anticipatory effects is not unprecedented. Some studies report early gender-based prediction in both L1 and L2 comprehension, though with temporal delays in L2 comprehension (e.g., Bosch & Foppolo, 2022; Dussias et al., 2013; Hopp, 2013, 2018), while others have failed to observe such effects in second language speakers (Lew-Williams & Fernald, 2010). I do not rule out that the reason for the lack of prediction in the current paradigm is a short determiner–noun interval (mean = 450 ms), which leaves limited time for pre-activation to manifest in eye movements. By contrast, studies that inserted longer pauses or intervening material often found anticipatory effects (e.g., Bosch & Foppolo, 2022; Huettig & Brouwer, 2015). Thus, the present results closely align with accounts that emphasize the time- and resource-dependent nature of prediction (Ito et al., 2016; Pickering & Garrod, 2013). Morphosyntactic features may be pre-activated, but the window for detecting them behaviorally is narrow.

Instead, facilitation effects appeared post-onset for the neurotypical L1 and L2 groups, who oriented their eye gaze more rapidly to the target noun once it was heard in the presence of a disambiguating determiner. These findings suggest that morphosyntactic cues may aid lexical recognition. That is, determiners likely narrowed down the set of possible continuations in advance, even if anticipatory looks were not captured. This interpretation resonates with ERP findings where gender-mismatched articles or adjectives elicit enhanced negativities relative to gender-matched ones. Such effects have been taken as evidence that comprehenders pre-activate morphosyntactic features of an upcoming noun, such that violations of these predictions trigger increased processing costs (e.g., Foucart et al., 2014; Ito et al., 2020; Otten & Van Berkum, 2009; Otten et al., 2007; Wicha et al., 2004). However, it is important to note that many ERP paradigms assess responses after word onset, leaving open whether the observed effects truly reflect pre-activation or rather facilitated post-lexical integration. As Pulvermüller and Grisoni (2020) note, ERP responses observed several hundred milliseconds after stimulus onset are difficult to interpret as unequivocal evidence of prediction, since prediction by definition must occur before the critical input is encountered. Instead, N400 reductions may be more conservatively interpreted as reflecting the consequences of prediction, such as faster access or easier integration, rather than unambiguous proof of anticipatory activation. In this respect, the post-target facilitation effects observed in the present study parallel the logic of much ERP work: they demonstrate that gender cues shaped subsequent word recognition, even if no anticipatory eye movements were visible before target onset. Importantly, L2 participants benefited from these cues to the same temporal extent as L1 speakers, indicating efficient use of morphosyntactic information in real time. At the same time, similar to what I observed during contextual prediction in Experiment 1a, their gaze distributions were more diffuse, consistent with broader evidence of less selective allocation of attention in L2 comprehension due to reduced certainty (Corps et al., 2023; Mitsugi, 2020).

By contrast, individuals with dyslexia showed no facilitation. This finding is particularly striking given the extensive evidence reviewed in Section 1.4.2, which showed that individuals with dyslexia struggle not only with offline grammatical tasks but also with real-time comprehension of structures involving long-distance dependencies. The present results demonstrate that these difficulties extend to the rapid integration of grammatical gender cues during spoken language processing, suggesting that dyslexia affects the ability to exploit morphosyntactic information incrementally as it unfolds in the speech signal. This suggests that prediction delays in dyslexia are not confined to semantic competitor activation, as observed in Experiment 2, or contextual anticipation in lower probability contexts (Engelhardt et al., 2021), but extend to morphosyntactic cues as well, supporting findings by Huettig and Brouwer (2015). One possibility is that the short determiner-noun interval particularly disadvantaged this group, whose predictive processing has been shown to emerge only under supportive conditions. Alternatively, reduced verbal working memory, a cognitive factor consistently associated with dyslexia (Reis et al., 2020; Smith-Spark et al., 2003), has been proposed as a predictor of predictive processing efficiency in neurotypical adults: Huettig and Janse (2016) found that higher working memory capacity predicted more efficient anticipatory eye movements during grammatical gender prediction in Dutch, while Li and Qu (2023) demonstrated that high-span participants initiated semantic prediction approximately 500 ms earlier than low-span participants in Mandarin comprehension. However, it should be noted that working memory accounts are most compelling when information must be maintained over longer processing windows. Given the short determiner–noun interval of the present paradigm, it is less clear that working memory capacity was the primary limiting factor here. More broadly, whether the absent facilitation reflects a deficit at the level of morphosyntactic representations, limitations in phonological encoding of agreement cues, or an interaction between the two remains an open question.

Finally, the lack of cross-linguistic congruency effects in Italian L2 speakers suggests that grammatical gender information from the Tyrolean L1 did not modulate online processing in this task. One possibility is that the mixed design with both congruent and incongruent items led participants to learn that cross-linguistic gender cues were not consistently reliable, and therefore to suppress. While speculative, this interpretation would suggest that predictive processing is sensitive to cue reliability and may be strategically reduced when cross-language overlap is inconsistent. However, because anticipatory use of gender cues was not observed even in the L1 group, the null effect cannot be taken as conclusive evidence that cross-linguistic morphosyntactic prediction is absent.

5.3 Limitations and Future Directions

5.3.1 Limitations

Several limitations of the present work should be acknowledged. First, the two main experiments differed in a number of **design features**, such as target position, syntactic variability, preview time, and task demands. These design differences were not deliberate manipulations, but stem from the fact that Experiment 2 closely replicated an existing paradigm (Ito & Husband, 2017). The goal was not to create identical experiments for direct statistical comparison, but rather to test similar predictive mechanisms across different linguistic environments. A detailed discussion of methodological differences is provided in Section 5.1.

A further consideration concerns the **short determiner–noun interval** in Experiment 1b. As discussed in Section 3.3, prior studies reporting anticipatory gender effects typically extended this interval through intervening adjectives or prosodic manipulations, which Italian sentences offer fewer opportunities to do without introducing additional grammatical gender cues. The absence of pre-target effects may therefore partly reflect the limited time available for anticipatory processing rather than an absence of morphosyntactic prediction per se. That said, this cannot fully account for all null findings: gender cues were successfully integrated post-onset by neurotypical L1 and L2 participants in the same paradigm, and the absence of such facilitation in the dyslexia group is more consistent with a processing-related explanation than with a timing artifact alone.

Another methodological concern relates to the **scoring of the word/non-word reading tasks** in Experiment 1. Italian word and pseudoword reading were evaluated by the author (a non-native Italian speaker trained by a native), whereas German word reading was scored by a native speaker of Standard German following standardized test guidelines. Nevertheless, participants with dyslexia consistently scored lower than typical L1 and L2 readers, in both accuracy and speed, suggesting that the group differences were captured reliably. Future work could benefit from independent coding by multiple raters to establish higher reliability. In Experiment 2, by contrast, the English reading tasks were scored by a native English speaker according to standardized criteria.

The **composition of the dyslexia sample** also imposes some limits. In Experiment 1, the dyslexia group was small, consisting of only 12 participants, which constrained statistical power. This limitation was addressed in Experiment 2, where the dyslexia sample was larger and thus provided a more reliable test of predictive processing. Moreover, in both experiments, the participants with dyslexia were university students and therefore likely represented a relatively compensated subgroup of the dyslexic population. It remains unclear whether the same predictive patterns would generalise to individuals with more severe or uncompensated reading difficulties.

A further limitation concerns the **language of instructions** across experiments. In both studies, all instructions were presented in the test language (Italian in Experiment 1, English in Experiment 2), both in written form on the screen and orally by the experimenter. This ensured task comprehension for all participants, but for L2 speakers it may also have promoted activation of the test language prior to task onset, consistent with the concept of "language activation mode" in bilinguals (Grosjean, 2001). I was not in a position to provide instructions in participants' other language(s), which constrained this methodological choice. While such activation mode effects cannot be ruled out, it is important that the mode of presentation was consistent within each group. While recent evidence indicates that cross-language prediction effects in bilinguals are strongly modulated by experimental language context and emerge primarily under conditions of high lexical predictability (Yin & Pickering, 2026), the relevance of language-mode effects may differ across experimental designs. In Experiment 1b, which investigated cross-linguistic grammatical gender congruency, the critical sentences were non-predictive with respect to upcoming lexical content. Under such conditions, the extent to which cross-language activation contributes to anticipatory gender processing is unclear. Thus, future work could directly manipulate language context (e.g., one-language vs. two-language instruction) under otherwise identical non-predictive conditions to more explicitly test the interaction between contextual language activation and grammatical prediction.

A further consideration concerns **the typological profiles of the languages** tested. Engelhardt et al. (2024) argue that speakers of verb-medial (SVO) languages (e.g., English) may rely more strongly on predictive processing strategies than speakers of verb-final (SOV) languages (e.g., Japanese). This is because, in SVO languages, the verb appears earlier in the sentence and provides information about the event being described, including which types of arguments are likely to follow, allowing comprehenders to form more specific expectations about upcoming input. In contrast, in verb-final languages, this information becomes available only at the end of the clause, which may reduce opportunities for early prediction and increase reliance on maintaining information in memory until integration can occur. While Italian is a canonical SVO language, Tyrolean, spoken by the L2 group, is genealogically related to Bavarian — a southern Germanic variety analysed as exhibiting SOV base order with verb-movement to second position in main clauses (V2) (see Bayer, 1984; Grewendorf & Weiss, 2014, on Bavarian syntax). This typological asymmetry could, in principle, influence comprehension strategies between the two groups. However, Engelhardt et al. acknowledge that languages which do not fall neatly into either the SVO or SOV category present a more complex picture, where the relationship between word order and predictive processing is less straightforward (see also R. P. Levy & Keller, 2013). In addition, it remains unclear how experience with a second language, particularly one with a different typological profile, may modulate prediction strategies in the native language, which is directly relevant to the present study, given that the Tyrolean-speaking participants are also proficient Italian speakers.

Finally, the operationalisation of **phonological similarity** was simplified.

Although phonological overlap was computed on the basis of broad phonetic transcriptions of the recorded utterances, ensuring alignment with the acoustic signal perceived by listeners, competitors were still defined using a string-based proportion of overlap on symbolic phonemic representations (Simmons & Magnuson, 2018). While this method aligns with prior visual-world studies, it treats phonemes as categorical symbols and therefore does not capture graded phonological similarity or phonetic/acoustic distance. Other approaches exist that estimate similarity from articulatory features or from how sounds pattern in actual speech (e.g., variation across dialects or acoustic measurements) (e.g., Kondrak, 2003; Wieling et al., 2012). Future work could incorporate feature-weighted or acoustically grounded measures to examine whether finer-grained similarity metrics explain additional variance in fixation behavior.

5.3.2 Future Directions

The present work opens several avenues for further investigation. First, future studies should systematically manipulate contextual constraint. Both experiments presented in this work relied primarily on high-cloze materials. While such contexts maximize the likelihood of observing anticipatory effects on the predicted target, as we have observed, they may paradoxically reduce evidence of competition, because a highly predictable target can dominate processing and suppress activation of alternative candidates. To fully capture the dynamics of semantic and phonological competition, future studies should extend these paradigms to lower-cloze contexts, where target predictions are weaker and competitors may have more opportunity to surface. In such designs, it would be informative to introduce graded phonological overlap conditions analogous to the Cp+Sem versus Cp-only manipulation in Experiment 2. For example, one competitor could overlap only with the predicted target, while another could overlap both with the target and with an additional contextually licensed word. This would allow researchers to test whether phonological activation remains confined to the exact-match form, or whether it can spread more flexibly when reinforced by multiple links.

Moreover, future work could adopt more precise measures of phonological similarity, which go beyond categorical overlap and are sensitive to graded phonetic or acoustic distance (e.g., Kondrak, 2003; Wieling et al., 2012). A further extension would be to compare onset-based competitors with rhyme-based competitors, to determine whether prediction engages different types of phonological overlap.

Another promising direction would be to adapt reaction-time paradigms that parallel production evidence but do not require articulation. For instance, lexical decision tasks following constraining sentence contexts could test whether facilitation emerges only for exact matches or also for phonological neighbors. Such methods would provide a more direct comprehension analogue to production findings (Drake & Corley, 2014) and could help adjudicate between lemma-only and suppression accounts of phonological prediction.

The lack of cross-linguistic grammatical gender activation found in Experiment 1b highlights the need to examine how reliability and consistency of cues shape predictive strategies. It would be informative to compare purely congruent contexts, purely incongruent contexts, and mixed contexts (like the one tested here). Such contrasts could test whether the system adapts its reliance on cross-linguistic cues according to their contextual utility, and whether adaptability differs by proficiency or dominance profiles.

5.4 Conclusion

The aim of this dissertation was to investigate how predictive mechanisms operate in spoken language comprehension across semantic, phonological, and morphosyntactic levels, and how these mechanisms vary in adults with dyslexia and in second-language speakers. To my knowledge, this is the first work to examine all three representational levels while including both L2 speakers and adults with dyslexia alongside typical L1 readers. Although the experiments differed somewhat in their design, this variation is not a limitation but a strength: predictive effects were observed across tasks and contexts, highlighting their robustness rather than being tied to a single paradigm.

By examining adults with dyslexia and L2 speakers, this dissertation was able to address questions about prediction that cannot be answered through typical L1 populations alone. Both groups engaged prediction but showed reduced selectivity. For L2 speakers, reduced selectivity may reflect weaker or less entrenched lexical–semantic links in the L2, a broader reliance on surface-level cues when comprehending in their second language, or the structural properties of the L2 lexicon itself. In both experiments, there was no consistent evidence for group differences in onset timing of target preactivation. In Experiment 1, L2 participants showed a later divergence point (approximately 250 ms), whereas no such timing difference was observed in Experiment 2. This discrepancy may be related to language dominance, with L2 speakers in Experiment 1 reporting greater daily use of their L1. For individuals with dyslexia, reduced selectivity may arise from noisier phonological representations or broader cognitive resource limitations. In both groups, prediction was present but less efficient, with greater uncertainty in narrowing down alternatives and measurable differences in either timing or fixation proportions compared to L1 typical readers. There is still an open question of why these group differences occur. Addressing this will require future research, ideally combining behavioral methods with neuroimaging and other approaches to uncover the neural mechanisms and resource demands that shape predictive processing.

Beyond group comparisons, the present findings contribute to a more refined understanding of semantic prediction as a psycholinguistic phenomenon. Although prediction was consistently initiated by highly constraining semantic contexts, its spread beyond the target word differed across representational levels. Anticipatory effects extended reliably to competitors that were semantically related to the target and supported by the sentence context (CP+Sem), whereas prediction did not consistently propagate to phonological cohorts or to target-related competitors that lacked contextual support (CP-only). This CP+Sem/CP-only dissociation shows that semantic pre-activation is not a product of indiscriminate spreading activation from the predicted word. Instead, it is selective: only competitors that are independently supported by the sentence context achieve above-baseline activation. This selectivity has direct implications for theoretical models of predictive processing. It argues against prediction as simply propagating outward from the most activated lexical item, and supports constraint-based views in which multiple contextual sources converge to license activation.

Another contribution concerns how the prediction system operates more broadly. The combined pattern across semantic, phonological, and morphosyntactic levels suggests that predictive processing is not only selective but also efficiency-driven. Rather than activating all related representations, the system appears to prioritise those that reduce uncertainty in the current context, while suppressing or bypassing alternatives that would not contribute to successful interpretation. This perspective also offers a reinterpretation of null phonological effects under high constraint: rather than indicating that prediction failed to reach the phonological level, such absences may reflect the system operating efficiently, suppressing activation of alternatives that would not reduce contextual uncertainty and impose unnecessary processing costs. This dissertation thus advances theoretical accounts of prediction as adaptive, context-sensitive, and efficient, while providing new evidence from under-studied groups that extends these accounts beyond the typical L1 population.

Bibliography

Acheson, D. J., Wells, J. B., & MacDonald, M. C. (2008). New and updated tests of print exposure and reading abilities in college students [Place: US]. *Behavior Research Methods*, *40*(1), 278–289. https://doi.org/10.3758/BRM.40.1.278

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, *38*(4), 419–439. https://doi.org/10.1006/jmla.1997.2558

AlpiLinK. (2023). Tyrolean. Retrieved August 7, 2025, from https://alpilink.it/en/tirolese/

Altmann, G. T. M., & Mirković, J. (2009). Incrementality and prediction in human sentence processing. *Cognitive Science*, *33*(4), 583–609. https://doi.org/10.1111/j.1551-6709.2009.01022.x

Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*(3), 247–264. https://doi.org/10.1016/S0010-0277(99)00059-1

Arai, M., & Keller, F. (2013). The use of verb-specific information for prediction in sentence processing. *Language and Cognitive Processes*, *28*(4), 525–560. https://doi.org/10.1080/01690965.2012.658072

Araújo, S., Inácio, F., Francisco, A., Faísca, L., Petersson, K. M., & Reis, A. (2011). Component processes subserving rapid automatized naming in dyslexic and non-dyslexic readers. *Dyslexia*, *17*(3), 242–255. https://doi.org/10.1002/dys.433

Armon-Lotem, S., & Marinis, T. (2015, December). Sentence Repetition. In S. Armon-Lotem, J. De Jong, & N. Meir (Eds.), *Assessing Multilingual Children* (pp. 95–122). Multilingual Matters. https://doi.org/10.21832/9781783093137-007

Arosio, F., Pagliarini, E., Perugini, M., Barbieri, L., & Guasti, M. T. (2016). Morphosyntax and logical abilities in Italian poor readers: The problem of SLI under-identification. *First Language*, *36*(3), 295–315. https://doi.org/10.1177/0142723716639501

ASTAT. (2024). *Astat-Info 56. Risultati censimento linguistico 2024* (Census Results). Provincia Autonoma di Bolzano-Alto Adige. Bolzano. https://astat.provincia.bz.it/it/pubblicazioni/risultati-censimento-linguistico-2024

Audacity Team. (2024). Audacity(R). https://www.audacityteam.org

Aveni, K., Ahmed, J., Borovsky, A., McRae, K., Jenkins, M. E., Sprengel, K., Fraser, J. A., Orange, J. B., Knowles, T., & Roberts, A. C. (2023). Predictive language comprehension in Parkinson's disease (D. Mirman, Ed.). *PLOS ONE*, *18*(2), e0262504. https://doi.org/10.1371/journal.pone.0262504

Balota, D. A. (1990). The role of meaning in word recognition. In *Comprehension processes in reading* (pp. 9–32). Lawrence Erlbaum Associates, Inc.

Bannon, J., Gollan, T. H., & Ferreira, V. S. (2025). Is predicting during language processing worth it? Effects of cloze probability and semantic similarity on failed predictions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *51*(1), 106–118. https://doi.org/10.1037/xlm0001347

Bar, M. (2007). The proactive brain: Using analogies and associations to generate predictions. *Trends in Cognitive Sciences*, *11*(7), 280–289. https://doi.org/10.1016/j.tics.2007.05.005

Barbosa, T., Rodrigues, C. C., Mello, C. B. D., Silva, M. C. D. S. E., & Bueno, O. F. A. (2019). Executive functions in children with dyslexia. *Arquivos de Neuro-Psiquiatria*, *77*(4), 254–259. https://doi.org/10.1590/0004-282x20190033

Barr, D. J. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, *59*(4), 457–474. https://doi.org/10.1016/j.jml.2007.09.002

Bayer, J. (1984). Comp in Bavarian Syntax. *The Linguistic Review*, *3*(3). https://doi.org/10.1515/tlir.1984.3.3.209

Belsley, D. A., Kuh, E., & Welsch, R. E. (1980, June). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity* (1st ed.). Wiley. https://doi.org/10.1002/0471725153

Boets, B., Op De Beeck, H. P., Vandermosten, M., Scott, S. K., Gillebert, C. R., Mantini, D., Bulthé, J., Sunaert, S., Wouters, J., & Ghesquière, P. (2013). Intact But Less Accessible Phonetic Representations in Adults with Dyslexia. *Science*, *342*(6163), 1251–1254. https://doi.org/10.1126/science.1244333

Borovsky, A., Elman, J. L., & Fernald, A. (2012). Knowing a lot for one's age: Vocabulary skill and not age is associated with anticipatory incremental sentence interpretation in children and adults. *Journal of Experimental Child Psychology*, *112*(4), 417–436. https://doi.org/10.1016/j.jecp.2012.01.005

Bosch, J. E., & Foppolo, F. (2022). Predictive processing of grammatical gender in bilingual children: The effect of cross-linguistic incongruency and language dominance. *Lingue e linguaggio*, (1), 5–27. https://doi.org/10.1418/104447

Boudewyn, M. A., Long, D. L., & Swaab, T. Y. (2015). Graded expectations: Predictive processing and the adjustment of expectations during spoken language comprehension. *Cognitive, Affective, & Behavioral Neuroscience*, *15*(3), 607–624. https://doi.org/10.3758/s13415-015-0340-0

Bovolenta, G., & Marsden, E. (2022). Prediction and Error-Based Learning in L2 Processing and Acquisition: A Conceptual Review. *Studies in Second Language Acquisition*, *44*(5), 1384–1409. https://doi.org/10.1017/S0272263121000723

Bovolenta, G., & Marsden, E. (2024). The effect of verb surprisal on the acquisition of second language syntactic structures in adults: An artificial language learning study. *Applied Psycholinguistics*, *45*(1), 110–137. https://doi.org/10.1017/S0142716423000334

Brand, J., Diamond, S. G., Thomas, N., & Gilbert-Diamond, D. (2021). Evaluating the data quality of the Gazepoint GP3 low-cost eye tracker when used independently by study participants. *Behavior Research Methods*, (53), 1502–1514. https://doi.org/https://doi.org/10.3758/s13428-020-01504-2

Branzi, F., Chang, Y.-N., Gaele, C., & Alexopoulou, T. (2021, November). The impact of L2 proficiency on cross-language influence during L2 word processing. https://doi.org/10.33774/coe-2021-fv5gt

Brothers, T., Morgan, E., Yacovone, A., & Kuperberg, G. (2023). Multiple predictions during language comprehension: Friends, foes, or indifferent companions? *Cognition*, *241*, 105602. https://doi.org/10.1016/j.cognition.2023.105602

Brothers, T., Swaab, T. Y., & Traxler, M. J. (2017). Goals and strategies influence lexical prediction during sentence comprehension. *Journal of Memory and Language*, *93*, 203–216. https://doi.org/10.1016/j.jml.2016.10.002

Brouwer, S., Özkan, D., & Küntay, A. C. (2019). Verb-based prediction during language processing: The case of Dutch and Turkish. *Journal of Child Language*, *46*(1), 80–97. https://doi.org/10.1017/S0305000918000375

Bui, G. (2023). A Dual-Motivation System in L2 and L3 Learning: A Theoretical Framework and Pedagogical Application. *Languages*, *8*(1), 69. https://doi.org/10.3390/languages8010069

Cardinaletti, A., Piccoli, E., & Volpato, F. (2022, December). Dyslexia and Syntactic Deficits: Overview and a Case Study of Language Training. In G. Cappelli & S. Noccetti (Eds.), *A Linguistic Approach to the Study of Dyslexia* (pp. 188–210). Multilingual Matters. https://doi.org/10.21832/9781800415973-010

Cardinaletti, A., & Volpato, F. (2015, August). On the comprehension and production of passive sentences and relative clauses by Italian university students with dyslexia. In E. Di Domenico, C. Hamann, &

S. Matteini (Eds.), *Linguistik Aktuell/Linguistics Today* (pp. 279–302, Vol. 223). John Benjamins Publishing Company. https://doi.org/10.1075/la.223.13car

Carioti, D., Masia, M. F., Travellini, S., & Berlingeri, M. (2021). Orthographic depth and developmental dyslexia: A meta-analytic study. *Annals of Dyslexia*, *71*(3), 399–438. https://doi.org/10.1007/s11881-021-00226-0

Carroll, D. W. (2008). *Psychology of language* (5th ed). Thomson/Wadsworth.

Casalicchio, J. (2024). *Dialetti d'Italia*. Carocci editore.

Cenoz, J., Hufeisen, B., & Jessner, U. (Eds.). (2001). *Cross-linguistic influence in third language aquisition: Psycholinguistic perspectives*. Multilingual Matters.

Chambers, C. G., & Cooke, H. (2009). Lexical competition during second-language listening: Sentence context, but not proficiency, constrains interference from the native lexicon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(4), 1029–1040. https://doi.org/10.1037/a0015901

Chambers, C. G., Tanenhaus, M. K., & Magnuson, J. S. (2004). Actions and Affordances in Syntactic Ambiguity Resolution. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(3), 687–696. https://doi.org/10.1037/0278-7393.30.3.687

Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review*, *113*(2), 234–272. https://doi.org/10.1037/0033-295X.113.2.234

Ciccolone, S. (2010). *Lo standard tedesco in Alto Adige: L'orientamento alla norma dei tedescofoni sudtirolesi*. LED Edizioni Universitarie di Lettere Economia Diritto.

Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics*, *27*(1), 3–42. https://doi.org/10.1017/S0142716406060024

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204. https://doi.org/10.1017/S0140525X12000477

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*(6), 407–428. https://doi.org/10.1037/0033-295X.82.6.407

Comrie, B., Haspelmath, M., & Bickel, B. (2015, May). *The Leipzig Glossing Rules: Conventions for interlinear morpheme-by-morpheme glosses* (tech. rep.). Max Planck Institute for Evolutionary Anthropology & University of Leipzig, Department of Linguistics. https://www.eva.mpg.de/lingua/pdf/Glossing-Rules.pdf

Connell, K., Puscama, M., Pinzon-Coimbra, J., Rembalsky, J., Xu, G., & Valdés Kroff, J. (2021). Phonologically Cued Lexical Anticipation in L2 English: A Visual World Eye-Tracking Study.

Cook, S., Pandža, N. B., Lancaster, A. K., & Gor, K. (2016). Fuzzy Nonnative Phonolexical Representations Lead to Fuzzy Form-to-Meaning Mappings. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.01345

Cook, V. (1999). Going beyond the Native Speaker in Language Teaching. *TESOL Quarterly*, *33*(2), 185. https://doi.org/10.2307/3587717

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language. *Cognitive Psychology*, *6*(1), 84–107. https://doi.org/10.1016/0010-0285(74)90005-X

Corps, R. E., Liao, M., & Pickering, M. J. (2023). Evidence for two stages of prediction in non-native speakers: A visual-world eye-tracking study. *Bilingualism: Language and Cognition*, *26*(1), 231–243. https://doi.org/10.1017/S1366728922000499

Costa, A., Caramazza, A., & Sebastian-Galles, N. (2000). The Cognate Facilitation Effect: Implications for Models of Lexical Access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(5), 1283–1296.

Cunningham, A. E., & Stanovich, K. E. (1991). Tracking the unique effects of print exposure in children: Associations with vocabulary, general knowledge, and spelling. *Journal of Educational Psychology*, *83*(2), 264–274. https://doi.org/10.1037/0022-0663.83.2.264

Dahan, D., & Magnuson, J. S. (2006). Spoken Word Recognition. In *Handbook of Psycholinguistics* (pp. 249–283). Elsevier. https://doi.org/10.1016/B978-012369374-7/50009-2

Dal Negro, S. (2010). Bilinguismo e diglossia. https://www.treccani.it/enciclopedia/bilinguismo-e-diglossia_(Enciclopedia-dell%27Italiano)/

Dal Negro, S. (2017). Bilinguismo asimmetrico in Alto Adige: Lo spazio sociolinguistico dell'italiano. In *Nuovi spazi comunicativi per l'italiano nel mondo: L'esperienza di Valori identitari e imprenditorialità.* (pp. 59–67).

Dal Negro, S. (2023). German and Romance varieties in contact in northeastern Italy. *Globe: A Journal of Language, Culture and Communication*, *15*, 80–92. https://doi.org/10.54337/ojs.globe.v15i.8039

Daneman, M., & Carpenter, P. A. (1983). Individual differences in integrating information between and within sentences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*(4), 561–584. https://doi.org/10.1037/0278-7393.9.4.561

Dauer, R. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics*, *11*(1), 51–62. https://doi.org/10.1016/S0095-4470(19)30776-4

Davies, A. (2003). *The native speaker: Myth and reality* (2nd ed). Multilingual Matters.

De Angelis, G. (2007). *Third or additional language acquisition.* Multilingual Matters.

De Groot, A. M. B., & Christoffels, I. K. (2006). Language control in bilinguals: Monolingual tasks and simultaneous interpreting. *Bilingualism: Language and Cognition*, *9*(2), 189–201. https://doi.org/10.1017/S1366728906002537

De Groot, A. M. B., & Keijzer, R. (2000). What Is Hard to Learn Is Easy to Forget: The Roles of Word Concreteness, Cognate Status, and Word Frequency in Foreign-Language Vocabulary Learning and Forgetting.

Language Learning, *50*(1), 1–56. https://doi.org/10.1111/0023-8333.00110
Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*(3), 283–321. https://doi.org/10.1037/0033-295X.93.3.283
Dell, G. S., & Chang, F. (2014). The P-chain: Relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1634), 20120394. https://doi.org/10.1098/rstb.2012.0394
DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, *8*(8), 1117–1121. https://doi.org/10.1038/nn1504
Denckla, M. B., & Rudel, R. G. (1976). Rapid 'automatized' naming (R.A.N.): Dyslexia differentiated from other learning disabilities. *Neuropsychologia*, *14*(4), 471–479. https://doi.org/10.1016/0028-3932(76)90075-0
Desroches, A. S., Joanisse, M. F., & Robertson, E. K. (2006). Specific phonological impairments in dyslexia revealed by eyetracking. *Cognition*, *100*(3), B32–B42. https://doi.org/10.1016/j.cognition.2005.09.001
Dijkgraaf, A., Hartsuiker, R. J., & Duyck, W. (2017). Predicting upcoming information in native-language and non-native-language auditory word recognition. *Bilingualism: Language and Cognition*, *20*(5), 917–930. https://doi.org/10.1017/S1366728916000547
Dijkgraaf, A., Hartsuiker, R. J., & Duyck, W. (2019). Prediction and integration of semantics during L2 and L1 listening. *Language, Cognition and Neuroscience*, *34*(7), 881–900. https://doi.org/10.1080/23273798.2019.1591469
Dijkstra, T., & Van Heuven, W. J. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, *5*(3), 175–197. https://doi.org/10.1017/S1366728902003012

Dink, J., & Ferguson, B. (2015). eyetrackingR: An R Library for Eye-tracking Data Analysis. Retrieved April 10, 2025, from http://www.eyetracking-r.com/

Dornic, S., & Ekehammar, B. (1988). Note on Self-Rated Second-Language Proficiency. *Perceptual and Motor Skills*, *66*(1), 40–42. https://doi.org/10.2466/pms.1988.66.1.40

Drake, E., & Corley, M. (2014). Effects in production of word pre-activation during listening: Are listener-generated predictions specified at a speech-sound level? *Memory & Cognition*, *43*(1), 111–120. https://doi.org/10.3758/s13421-014-0451-9

Dunn, L. M., & Dunn, D. M. (2007). Peabody Picture Vocabulary Test–Fourth Edition [Institution: American Psychological Association]. https://doi.org/10.1037/t15144-000

Dussias, P. E., Valdés Kroff, J. R., Guzzardo Tamargo, R. E., & Gerfen, C. (2013). When gender and looking go hand in hand: Grammatical gender processing in L2 Spanish. *Studies in Second Language Acquisition*, *35*(2), 353–387. https://doi.org/10.1017/S0272263112000915

Duyck, W. (2005). Translation and Associative Priming With Cross-Lingual Pseudohomophones: Evidence for Nonselective Phonological Activation in Bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(6), 1340–1359. https://doi.org/10.1037/0278-7393.31.6.1340

Engelhardt, P. E., Filipović, L., & Hawkins, J. A. (2024). Prediction in SVO and SOV languages: Processing and typological considerations. *Linguistics*, *62*(2), 349–383. https://doi.org/10.1515/ling-2022-0055

Engelhardt, P. E., Yuen, M. K. Y., Kenning, E. A., & Filipovic, L. (2021). Are Linguistic Prediction Deficits Characteristic of Adults with Dyslexia? *Brain Sciences*, *11*(1), 59. https://doi.org/10.3390/brainsci11010059

Favier, S., Meyer, A. S., & Huettig, F. (2021). Literacy can enhance syntactic prediction in spoken language processing. *Journal of Experimental Psychology: General*, *150*(10), 2167–2174. https://doi.org/10.1037/xge0001042

Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, *44*(4), 491–505. https://doi.org/10.1111/j.1469-8986.2007.00531.x

Federmeier, K. D., & Kutas, M. (1999). A Rose by Any Other Name: Long-Term Memory Structure and Sentence Processing. *Journal of Memory and Language*, *41*(4), 469–495. https://doi.org/10.1006/jmla.1999.2660

Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain Research*, *1146*, 75–84. https://doi.org/10.1016/j.brainres.2006.06.101

Fernandez, L. B., Engelhardt, P. E., Patarroyo, A. G., & Allen, S. E. (2020). Effects of speech rate on anticipatory eye movements in the visual world paradigm: Evidence from aging, native, and non-native language processing. *Quarterly Journal of Experimental Psychology*, *73*(12), 2348–2361. https://doi.org/10.1177/1747021820948019

Fernandez, L. B., Hadley, L. V., Gamboa, J. C., Allison, C., & Allen, S. E. (2025). The impact of speech rate on first- and second-stage prediction in L1 and L2 speakers. *Bilingualism: Language and Cognition*, 1–13. https://doi.org/10.1017/S1366728925100515

Fernandez, L. B., Shehzad, M., & Hadley, L. V. (2025). Younger adults may be faster at making semantic predictions, but older adults are more efficient. *Psychology and Aging*, *40*(3), 318–325. https://doi.org/10.1037/pag0000872

Ferreira, F., Bailey, K. G., & Ferraro, V. (2002). Good-Enough Representations in Language Comprehension. *Current Directions in Psychological Science*, *11*(1), 11–15. https://doi.org/10.1111/1467-8721.00158

Fine, A. B., Jaeger, T. F., Farmer, T. A., & Qian, T. (2013). Rapid Expectation Adaptation during Syntactic Comprehension. *PLoS ONE*, *8*(10), e77661. https://doi.org/10.1371/journal.pone.0077661

Forster, K. I. (1981). Priming and the Effects of Sentence and Lexical Contexts on Naming Time: Evidence for Autonomous Lexical

Processing. *The Quarterly Journal of Experimental Psychology Section A*, *33*(4), 465–495. https://doi.org/10.1080/14640748108400804

Foucart, A., Martin, C. D., Moreno, E. M., & Costa, A. (2014). Can bilinguals see it coming? Word anticipation in L2 sentence reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(5), 1461–1469. https://doi.org/10.1037/a0036756

Francis, W. S. (2020). Shared core meanings and shared associations in bilingual semantic memory: Evidence from research on implicit memory. *International Journal of Bilingualism*, *24*(3), 464–477. https://doi.org/10.1177/1367006918814375

Frazier, L. (1979). *On comprehending sentences: Syntactic parsing strategies* [Ph.D. dissertation]. University of Connecticut.

Frisson, S., Harvey, D. R., & Staub, A. (2017). No prediction error cost in reading: Evidence from eye movements. *Journal of Memory and Language*, *95*, 200–214. https://doi.org/10.1016/j.jml.2017.04.007

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138. https://doi.org/10.1038/nrn2787

Fromkin, V. A. (1971). The Non-Anomalous Nature of Anomalous Utterances. *Language*, *47*(1), 27–52. https://doi.org/10.2307/412187

Gardiner, I. A., & Deterding, D. (2025, March). Stress-Timing Versus Syllable-Timing in World Englishes. In K. Bolton (Ed.), *The Wiley Blackwell Encyclopedia of World Englishes* (1st ed., pp. 1–8). Wiley. https://doi.org/10.1002/9781119518297.eowe00334

Gazepoint. (2024). GP3 Eye-Tracker. https://www.gazept.com

Georgiou, G. K., Ghazyani, R., & Parrila, R. (2018). Are RAN deficits in university students with dyslexia due to defective lexical access, impaired anchoring, or slow articulation? *Annals of Dyslexia*, *68*(2), 85–103. https://doi.org/10.1007/s11881-018-0156-z

Giglio, L., Ostarek, M., Weber, K., & Hagoort, P. (2022). Commonalities and Asymmetries in the Neurobiological Infrastructure for Language

Production and Comprehension. *Cerebral Cortex*, *32*(7), 1405–1418. https://doi.org/10.1093/cercor/bhab287

Glazzard, J. (2010). The impact of dyslexia on pupils' self-esteem. *Support for Learning*, *25*(2), 63–69. https://doi.org/10.1111/j.1467-9604.2010.01442.x

Godfroid, A. (2020). *Eye tracking in second language acquisition and bilingualism: A research synthesis and methodological guide*. Routledge, Taylor & Francis Group. https://doi.org/10.4324/9781315775616

Gollan, T. H., Montoya, R. I., Cera, C., & Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language*, *58*(3), 787–814. https://doi.org/10.1016/j.jml.2007.07.001

Gollan, T. H., Montoya, R. I., Fennema-Notestine, C., & Morris, S. K. (2005). Bilingualism affects picture naming but not picture classification. *Memory & Cognition*, *33*(7), 1220–1234. https://doi.org/10.3758/BF03193224

Gor, K., Cook, S., Bordag, D., Chrabaszcz, A., & Opitz, A. (2021). Fuzzy Lexical Representations in Adult Second Language Speakers. *Frontiers in Psychology*, *12*, 732030. https://doi.org/10.3389/fpsyg.2021.732030

Grewendorf, G., & Weiss, H. (2014). *Bavarian syntax: Contributions to the theory of syntax*. John Benjamins Publishing Company.

Grosjean, F. (2001). The bilingual's language modes. In J. Nicol (Ed.), *One Mind, Two Languages: Bilingual Language Processing* (pp. 1–22). Blackwell Publishing.

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception & Psychophysics*, *28*(4), 267–283. https://doi.org/10.3758/BF03204386

Guasti, M. T., Branchini, C., Vernice, M., Barbieri, L., & Arosio, F. (2015, August). Language disorders in children with Developmental Dyslexia. In S. Stavrakaki (Ed.), *Language Acquisition and Language*

Disorders (pp. 35–56, Vol. 58). John Benjamins Publishing Company. https://doi.org/10.1075/lald.58.02gua

Haeuser, K. I., & Borovsky, A. (2024). Predictive processing suppresses form-related words with overlapping onsets. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *46*.

Hagoort, P. (2003). Interplay between Syntax and Semantics during Sentence Comprehension: ERP Effects of Combining Syntactic and Semantic Violations. *Journal of Cognitive Neuroscience*, *15*(6), 883–899. https://doi.org/10.1162/089892903322370807

Hale, J. (2001). A probabilistic earley parser as a psycholinguistic model. *Second meeting of the North American Chapter of the Association for Computational Linguistics on Language technologies 2001 - NAACL '01*, 1–8. https://doi.org/10.3115/1073336.1073357

Hammarberg, B. (2010). The languages of the multilingual: Some conceptual and terminological issues. *IRAL - International Review of Applied Linguistics in Language Teaching*, *48*(2-3). https://doi.org/10.1515/iral.2010.005

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*. Springer New York. https://doi.org/10.1007/978-0-387-84858-7

Hawkins, J. A. (2004, November). *Efficiency and Complexity in Grammars*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199252695.001.0001

Hawkins, J. A. (2014, February). *Cross-Linguistic Variation and Efficiency* (1st ed.). Oxford University PressOxford. https://doi.org/10.1093/acprof:oso/9780199664993.001.0001

Hayiou-Thomas, M. E., Carroll, J. M., & Snowling, M. J. (2018, August). Developmental Dyslexia. In S.-A. Rueschemeyer & M. G. Gaskell (Eds.), *The Oxford Handbook of Psycholinguistics* (pp. 783–813). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198786825.013.34

Heim, S., & Grande, M. (2012). Fingerprints of developmental dyslexia. *Trends in Neuroscience and Education*, *1*(1), 10–14. https://doi.org/10.1016/j.tine.2012.09.001

Heselwood, B. (2013). *Phonetic transcription in theory and practice*. Edinburgh University Press.

Hintz, F., Meyer, A. S., & Huettig, F. (2017). Predictors of verb-mediated anticipatory eye movements in the visual world. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(9), 1352–1374. https://doi.org/10.1037/xlm0000388

Hoeks, J. C. J., Stowe, L. A., & Doedens, G. (2004). Seeing words in context: The interaction of lexical and sentence level information during reading. *Cognitive Brain Research*, *19*(1), 59–73. https://doi.org/10.1016/j.cogbrainres.2003.10.022

Hopp, H. (2013). Grammatical gender in adult L2 acquisition: Relations between lexical and syntactic variability. *Second Language Research*, *29*(1), 33–56. https://doi.org/10.1177/0267658312461803

Hopp, H. (2015). Individual differences in the second language processing of object–subject ambiguities. *Applied Psycholinguistics*, *36*(2), 129–173. https://doi.org/10.1017/S0142716413000180

Hopp, H. (2018). The Bilingual Mental Lexicon in L2 Sentence Processing. *Second Language*, *17*, 5–27.

Hopp, H., & Lemmerth, N. (2018). Lexical and Syntactic Congruency in L2 Predictive Gender Processing. *Studies in Second Language Acquisition*, *40*(1), 171–199. https://doi.org/10.1017/S0272263116000437

Hoshino, N., & Kroll, J. F. (2008). Cognate effects in picture naming: Does cross-language activation survive a change of script? *Cognition*, *106*(1), 501–511. https://doi.org/10.1016/j.cognition.2007.02.001

Huettig, F. (2015). Four central questions about prediction in language processing. *Brain Research*, *1626*, 118–135. https://doi.org/10.1016/j.brainres.2015.02.014

Huettig, F. (2025, March). Defining Prediction. In *Looking Ahead: The New Science of the Predictive Mind* (1st ed., pp. 12–17). Cambridge University Press. https://doi.org/10.1017/9781009245470.004

Huettig, F., & Altmann, G. T. (2005). Word meaning and the control of eye fixation: Semantic competitor effects and the visual world paradigm. *Cognition*, *96*(1), B23–B32. https://doi.org/10.1016/j.cognition.2004.10.003

Huettig, F., Audring, J., & Jackendoff, R. (2022). A parallel architecture perspective on pre-activation and prediction in language processing. *Cognition*, *224*, 105050. https://doi.org/10.1016/j.cognition.2022.105050

Huettig, F., & Brouwer, S. (2015). Delayed Anticipatory Spoken Language Processing in Adults with Dyslexia-Evidence from Eye-tracking: Word Reading and Predictive Language Processing. *Dyslexia*, *21*(2), 97–122. https://doi.org/10.1002/dys.1497

Huettig, F., & Guerra, E. (2019). Effects of speech rate, preview time of visual context, and participant instructions reveal strong limits on prediction in language processing. *Brain Research*, *1706*, 196–208. https://doi.org/10.1016/j.brainres.2018.11.013

Huettig, F., & Janse, E. (2016). Individual differences in working memory and processing speed predict anticipatory spoken language processing in the visual world. *Language, Cognition and Neuroscience*, *31*(1), 80–93. https://doi.org/10.1080/23273798.2015.1047459

Huettig, F., & McQueen, J. M. (2007). The tug of war between phonological, semantic and shape information in language-mediated visual search. *Journal of Memory and Language*, *57*(4), 460–482. https://doi.org/10.1016/j.jml.2007.02.001

Huettig, F., & Mishra, R. K. (2014). How Literacy Acquisition Affects the Illiterate Mind – A Critical Examination of Theories and Evidence. *Language and Linguistics Compass*, *8*(10), 401–427. https://doi.org/10.1111/lnc3.12092

Huettig, F., & Pickering, M. J. (2019). Literacy Advantages Beyond Reading: Prediction of Spoken Language. *Trends in Cognitive Sciences*, *23*(6), 464–475. https://doi.org/10.1016/j.tics.2019.03.008

Humphrey, N., & Mullins, P. M. (2002). Self-concept and self-esteem in developmental dyslexia. *Journal of Research in Special Educational Needs*, *2*(2), j.1471–3802.2002.00163.x. https://doi.org/10.1111/j.1471-3802.2002.00163.x

Ito, A. (2019). Prediction of orthographic information during listening comprehension: A printed-word visual world study. *Quarterly Journal of Experimental Psychology*, *72*(11), 2584–2596. https://doi.org/10.1177/1747021819851394

Ito, A. (2024). Phonological prediction during comprehension: A review and meta-analysis of visual-world eye-tracking studies. *Journal of Memory and Language*, *139*, 104553. https://doi.org/10.1016/j.jml.2024.104553

Ito, A., Corley, M., Pickering, M. J., Martin, A. E., & Nieuwland, M. S. (2016). Predicting form and meaning: Evidence from brain potentials. *Journal of Memory and Language*, *86*, 157–171. https://doi.org/10.1016/j.jml.2015.10.007

Ito, A., Gambi, C., Pickering, M. J., Fuellenbach, K., & Husband, E. M. (2020). Prediction of phonological and gender information: An event-related potential study in Italian. *Neuropsychologia*, *136*, 107291. https://doi.org/10.1016/j.neuropsychologia.2019.107291

Ito, A., & Husband, E. M. (2017). How robust are effects of semantic and phonological prediction during language comprehension? A visual world eye-tracking study. *Technical Committee on Thought and Language*, *117*(149).

Ito, A., & Knoeferle, P. (2023). Analysing data from the psycholinguistic visual-world paradigm: Comparison of different analysis methods. *Behavior Research Methods*, *55*(7), 3461–3493. https://doi.org/10.3758/s13428-022-01969-3

Ito, A., Martin, A. E., & Nieuwland, M. S. (2017a). On predicting form and meaning in a second language. *Journal of Experimental Psychology:*

Learning, Memory, and Cognition, *43*(4), 635–652. https://doi.org/10.1037/xlm0000315
Ito, A., Martin, A. E., & Nieuwland, M. S. (2017b). How robust are prediction effects in language comprehension? Failure to replicate article-elicited N400 effects. *Language, Cognition and Neuroscience*, *32*(8), 954–965. https://doi.org/10.1080/23273798.2016.1242761
Ito, A., Nguyen, H. T. T., & Knoeferle, P. (2024). German-dominant Vietnamese heritage speakers use semantic constraints of German for anticipation during comprehension in Vietnamese. *Bilingualism: Language and Cognition*, *27*(1), 57–74. https://doi.org/10.1017/S136672892300041X
Ito, A., & Pickering, M. J. (2021, September). Chapter 2. Automaticity and prediction in non-native language comprehension. In E. Kaan & T. Grüter (Eds.), *Bilingual Processing and Acquisition* (pp. 26–46, Vol. 12). John Benjamins Publishing Company. https://doi.org/10.1075/bpa.12.02ito
Ito, A., Pickering, M. J., & Corley, M. (2018). Investigating the time-course of phonological prediction in native and non-native speakers of English: A visual world eye-tracking study. *Journal of Memory and Language*, *98*, 1–11. https://doi.org/10.1016/j.jml.2017.09.002
Ito, A., & Sakai, H. (2021). Everyday Language Exposure Shapes Prediction of Specific Words in Listening Comprehension: A Visual World Eye-Tracking Study. *Frontiers in Psychology*, *12*, 607474. https://doi.org/10.3389/fpsyg.2021.607474
Ivanova, I., & Costa, A. (2008). Does bilingualism hamper lexical access in speech production? *Acta Psychologica*, *127*(2), 277–288. https://doi.org/10.1016/j.actpsy.2007.06.003
Jackendoff, R. (2007). A Parallel Architecture perspective on language processing. *Brain Research*, *1146*, 2–22. https://doi.org/10.1016/j.brainres.2006.08.111
Jared, D., & Kroll, J. F. (2001). Do Bilinguals Activate Phonological Representations in One or Both of Their Languages When Naming

Words? *Journal of Memory and Language*, *44*(1), 2–31. https://doi.org/10.1006/jmla.2000.2747
Joanisse, M. F., Manis, F. R., Keating, P., & Seidenberg, M. S. (2000). Language Deficits in Dyslexic Children: Speech Perception, Phonology, and Morphology. *Journal of Experimental Child Psychology*, *77*(1), 30–60. https://doi.org/10.1006/jecp.1999.2553
Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99*(1), 122–149. https://doi.org/10.1037/0033-295X.99.1.122
Kaan, E. (2014). Predictive sentence processing in L2 and L1: What is different?*. *Linguistic Approaches to Bilingualism*, *4*(2), 257–282. https://doi.org/10.1075/lab.4.2.05kaa
Kaan, E., Kirkham, J., & Wijnen, F. (2016). Prediction and integration in native and second-language processing of elliptical structures. *Bilingualism: Language and Cognition*, *19*(1), 1–18. https://doi.org/10.1017/S1366728914000844
Kamide, Y., Altmann, G. T., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, *49*(1), 133–156. https://doi.org/10.1016/S0749-596X(03)00023-8
Kellerman, E. (1979). Transfer and Non-Transfer: Where We Are Now. *Studies in Second Language Acquisition*, *2*(1), 37–57. https://doi.org/10.1017/S0272263100000942
Kessler, R. C., Adler, L., Ames, M., Demler, O., Faraone, S., Hiripi, E., Howes, M. J., Jin, R., Secnik, K., Spencer, T., Ustun, T. B., & Walters, E. E. (2005). The World Health Organization Adult ADHD Self-Report Scale (ASRS): A short screening scale for use in the general population. *Psychological Medicine*, *35*(2), 245–256. https://doi.org/10.1017/s0033291704002892
Kim, J., Wessel, J. R., & Hendrickson, K. (2023). Inhibition of lexical representations after violated semantic predictions. *Cognition*, *240*, 105585. https://doi.org/10.1016/j.cognition.2023.105585

Knoeferle, P., & Crocker, M. W. (2006). The Coordinated Interplay of Scene, Utterance, and World Knowledge: Evidence From Eye Tracking. *Cognitive Science*, *30*(3), 481–529. https://doi.org/10.1207/s15516709cog0000_65

Knoeferle, P., & Crocker, M. W. (2007). The influence of recent scene events on spoken comprehension: Evidence from eye movements. *Journal of Memory and Language*, *57*(4), 519–543. https://doi.org/10.1016/j.jml.2007.01.003

Kochari, A. R., & Flecken, M. (2019). Lexical prediction in language comprehension: A replication study of grammatical gender effects in Dutch. *Language, Cognition and Neuroscience*, *34*(2), 239–253. https://doi.org/10.1080/23273798.2018.1524500

Kondrak, G. (2003). Phonetic Alignment and Similarity. *Computers and the Humanities*, *37*(3), 273–291. https://doi.org/10.1023/A:1025071200644

Kristina Moll & Karin Landerl. (2014). *SLRT-II - Lese- und Rechtschreibtest.* Normen. Retrieved October 26, 2023, from https://www.testzentrale.de/shop/lese-und-rechtschreibtest.html

Kukona, A., Braze, D., Johns, C. L., Mencl, W. E., Van Dyke, J. A., Magnuson, J. S., Pugh, K. R., Shankweiler, D. P., & Tabor, W. (2016). The real-time prediction and inhibition of linguistic outcomes: Effects of language and literacy skill. *Acta Psychologica*, *171*, 72–84. https://doi.org/10.1016/j.actpsy.2016.09.009

Kukona, A., Fang, S.-Y., Aicher, K. A., Chen, H., & Magnuson, J. S. (2011). The time course of anticipatory constraint integration. *Cognition*, *119*(1), 23–42. https://doi.org/10.1016/j.cognition.2010.12.002

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, *31*(1), 32–59. https://doi.org/10.1080/23273798.2015.1102299

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP). *Annual review of psychology*, *62*, 621. https://doi.org/10.1146/annurev.psych.093008.131123

Kutas, M., & Hillyard, S. A. (1980). Reading Senseless Sentences: Brain Potentials Reflect Semantic Incongruity. *Science*, *207*(4427), 203–205. https://doi.org/10.1126/science.7350657

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*(5947), 161–163. https://doi.org/10.1038/307161a0

Lai, M. K., Payne, B. R., & Federmeier, K. D. (2024). Graded and ungraded expectation patterns: Prediction dynamics during active comprehension. *Psychophysiology*, *61*(1), e14424. https://doi.org/10.1111/psyp.14424
e14424 PsyP-2022-0388.R2.

Landerl, K., Ramus, F., Moll, K., Lyytinen, H., Leppänen, P. H. T., Lohvansuu, K., O'Donovan, M., Williams, J., Bartling, J., Bruder, J., Kunze, S., Neuhoff, N., Tóth, D., Honbolygó, F., Csépe, V., Bogliotti, C., Iannuzzi, S., Chaix, Y., Démonet, J.-F., … Schulte-Körne, G. (2013). Predictors of developmental dyslexia in European orthographies with varying complexity. *Journal of Child Psychology and Psychiatry*, *54*(6), 686–694. https://doi.org/10.1111/jcpp.12029

Lanthaler, F. (2001). Zwischenregister der deutschen Sprache in Südtirol. In K. Egger & F. Lanthaler (Eds.), *Die deutsche Sprache in Südtirol: Einheitssprache und regionale Vielfalt* (pp. 137–152). Folio Verlag.
Cited in: Ciccolone, S. (2010). Lo standard tedesco in Alto Adige. LED Edizioni Universitarie.

Laufer, B., & Paribakht, T. S. (1998). The Relationship Between Passive and Active Vocabularies: Effects of LanguageLearning Context. *Language Learning*, *48*(3), 365–391. https://doi.org/10.1111/0023-8333.00046

Lefly, D. L., & Pennington, B. F. (2000). Reliability and validity of the adult reading history questionnaire. *Journal of Learning Disabilities*, *33*(3), 286–296. https://doi.org/10.1177/002221940003300306

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods*, *44*(2), 325–343. https://doi.org/10.3758/s13428-011-0146-0

Lenth, R., & Piaskowski, J. (2025). Emmeans: Estimated Marginal Means, aka Least-Squares Means. https://rvlenth.github.io/emmeans/

Leonardi, M. M. V. (2020). "I hardly ever practice the real Standard German." Self-reported language use and language proficiency in South Tyrol (Italy). *Linguistik Online*, *102*(2), 83–98. https://doi.org/10.13092/lo.102.6825

Leonardi, M. M. V. (2021). Attitudes of South Tyrolean University Students towards German Varieties. *Languages*, *6*(3), 137. https://doi.org/10.3390/languages6030137

Levelt, W. J. M. (1989, February). *Speaking: From Intention to Articulation*. The MIT Press. https://doi.org/10.7551/mitpress/6393.001.0001

Levy, B. J., McVeigh, N. D., Marful, A., & Anderson, M. C. (2007). Inhibiting Your Native Language: The Role of Retrieval-Induced Forgetting During Second-Language Acquisition. *Psychological Science*, *18*(1), 29–34. https://doi.org/10.1111/j.1467-9280.2007.01844.x

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177. https://doi.org/10.1016/j.cognition.2007.05.006

Levy, R. P., & Keller, F. (2013). Expectation and locality effects in German verb-final structures. *Journal of Memory and Language*, *68*(2), 199–222. https://doi.org/10.1016/j.jml.2012.02.005

Lew-Williams, C., & Fernald, A. (2010). Real-time processing of gender-marked articles by native and non-native Spanish speakers. *Journal of Memory and Language*, *63*(4), 447–464. https://doi.org/10.1016/j.jml.2010.07.003

Li, X., Li, X., & Qu, Q. (2022). Predicting phonology in language comprehension: Evidence from the visual world eye-tracking task in Mandarin Chinese. *Journal of Experimental Psychology: Human Perception and Performance*, *48*(5), 531–547. https://doi.org/10.1037/xhp0000999

Li, X., & Qu, Q. (2023). Verbal working memory capacity modulates semantic and phonological prediction in spoken comprehension. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-023-02348-5

Libben, M. R., & Titone, D. A. (2009). Bilingual lexical access in context: Evidence from eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(2), 381–390. https://doi.org/10.1037/a0014875

Linck, J. A., Kroll, J. F., & Sunderman, G. (2009). Losing Access to the Native Language While Immersed in a Second Language: Evidence for the Role of Inhibition in Second-Language Learning. *Psychological Science*, *20*(12), 1507–1515. https://doi.org/10.1111/j.1467-9280.2009.02480.x

Luce, P. A., & Pisoni, D. B. (1998). Recognizing Spoken Words: The Neighborhood Activation Model: *Ear and Hearing*, *19*(1), 1–36. https://doi.org/10.1097/00003446-199802000-00001

Luke, S. G., & Christianson, K. (2016). Limits on lexical prediction during reading. *Cognitive Psychology*, *88*, 22–60. https://doi.org/10.1016/j.cogpsych.2016.06.002

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

Ma, T., Chen, B., Lu, C., & Dunlap, S. (2015). Proficiency and sentence constraint effects on second language word learning. *Acta Psychologica*, *159*, 116–122. https://doi.org/10.1016/j.actpsy.2015.05.014

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, *101*(4), 676–703. https://doi.org/10.1037/0033-295X.101.4.676

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, *109*(2), 163–203. https://doi.org/10.1037/0033-2909.109.2.163

Magnuson, J. S., Mirman, D., Luthra, S., Strauss, T., & Harris, H. D. (2018). Interaction in Spoken Word Recognition Models: Feedback Helps. *Frontiers in Psychology*, *9*, 369. https://doi.org/10.3389/fpsyg.2018.00369

Magnuson, J. S., Tanenhaus, M. K., Aslin, R. N., & Dahan, D. (2003). The time course of spoken word learning and recognition: Studies with artificial lexicons. *Journal of Experimental Psychology: General*, *132*(2), 202–227. https://doi.org/10.1037/0096-3445.132.2.202

Mani, N., & Huettig, F. (2012). Prediction during language processing is a piece of cake — But only for skilled producers. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(4), 843–847. https://doi.org/10.1037/a0029284

Mani, N., & Huettig, F. (2014). Word reading skill predicts anticipation of upcoming spoken language input: A study of children developing proficiency in reading. *Journal of Experimental Child Psychology*, *126*, 264–279. https://doi.org/10.1016/j.jecp.2014.05.004

Marcus, D. J., Karatekin, C., & Markiewicz, S. (2006). Oculomotor evidence of sequence learning on the serial reaction time task. *Memory & Cognition*, *34*(2), 420–432. https://doi.org/10.3758/BF03193419

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). Language Experience and Proficiency Questionnaire (LEAP-Q). *Journal of Speech Language and Hearing Research*, *50*(4), 940–967.

Marian, V., Spivey, M., & Hirsch, J. (2003). Shared and separate systems in bilingual language processing: Converging evidence from eyetracking and brain imaging. *Brain and Language*, *86*(1), 70–82. https://doi.org/10.1016/S0093-934X(02)00535-7

Marotta, G. (2022, December). Developmental Dyslexia and Morphosyntactic Competence in Italian Young Adults. In G. Cappelli & S. Noccetti (Eds.), *A Linguistic Approach to the Study of Dyslexia* (pp. 165–187). Multilingual Matters. https://doi.org/10.21832/9781800415973-009

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, *25*(1-2), 71–102. https://doi.org/10.1016/0010-0277(87)90005-9

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive*

Psychology, *10*(1), 29–63. https://doi.org/10.1016/0010-0285(78)90018-X

Martin, C. D., Branzi, F. M., & Bar, M. (2018). Prediction is Production: The missing link between language production and comprehension. *Scientific Reports*, *8*(1), 1079. https://doi.org/10.1038/s41598-018-19499-4

Martin, C. D., Thierry, G., Kuipers, J.-R., Boutonnet, B., Foucart, A., & Costa, A. (2013). Bilinguals reading in their second language do not predict upcoming words as native readers do. *Journal of Memory and Language*, *69*(4), 574–588. https://doi.org/10.1016/j.jml.2013.08.001

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314–324. https://doi.org/10.3758/s13428-011-0168-7

Mcclelland, J., Mirman, D., & Holt, L. (2006). Are there interactive processes in speech perception? *Trends in Cognitive Sciences*, *10*(8), 363–369. https://doi.org/10.1016/j.tics.2006.06.007

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*(1), 1–86. https://doi.org/10.1016/0010-0285(86)90015-0

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review*, *88*(5), 375–407. https://doi.org/10.1037/0033-295X.88.5.375

McNulty, M. A. (2003). Dyslexia and the Life Course. *Journal of Learning Disabilities*, *36*(4), 363–381. https://doi.org/10.1177/00222194030360040701

McQueen, J. M., Dahan, D., & Cutler, A. (2003, January). Continuity and gradedness in speech processing. In N. O. Schiller & A. S. Meyer (Eds.), *Phonetics and Phonology in Language Comprehension and Production*. DE GRUYTER MOUTON. https://doi.org/10.1515/9783110895094.39

Meara, P. (1978). Learners' Word Associations in French. *Interlanguage Studies Bulletin*, *3*(2), 192–211.

Meara, P. (1983). Word associations in a foreign language. *Nottingham Linguistics Circular*, *11*, 29–38.

Melloni, C., & Vender, M. (2022). Morphological awareness in developmental dyslexia: Playing with nonwords in a morphologically rich language (C. Manouilidou, Ed.). *PLOS ONE*, *17*(11), e0276643. https://doi.org/10.1371/journal.pone.0276643

Mercier, J., Pivneva, I., & Titone, D. (2016). The role of prior language context on bilingual spoken word processing: Evidence from the visual world task. *Bilingualism: Language and Cognition*, *19*(2), 376–399. https://doi.org/10.1017/S1366728914000340

Mielke, J. (2012). A phonetically based metric of sound similarity. *Lingua*, *122*(2), 145–163. https://doi.org/10.1016/j.lingua.2011.04.006

Mishra, R. K., Singh, N., Pandey, A., & Huettig, F. (2012). Spoken language-mediated anticipatory eye-movements are modulated by reading ability - Evidence from Indian low and high literates. *Journal of Eye Movement Research*, *5*(1). https://doi.org/10.16910/jemr.5.1.3

Misra, M., Guo, T., Bobb, S. C., & Kroll, J. F. (2012). When bilinguals choose a single word to speak: Electrophysiological evidence for inhibition of the native language. *Journal of Memory and Language*, *67*(1), 224–237. https://doi.org/10.1016/j.jml.2012.05.001

Mitsugi, S. (2017). Incremental comprehension of Japanese passives: Evidence from the visual-world paradigm. *Applied Psycholinguistics*, *38*(4), 953–983. https://doi.org/10.1017/S0142716416000515

Mitsugi, S. (2020). Generating predictions based on semantic categories in a second language: A case of numeral classifiers in Japanese. *International Review of Applied Linguistics in Language Teaching*, *58*(3), 323–349. https://doi.org/10.1515/iral-2017-0118

Mitsugi, S., & Macwhinney, B. (2016). The use of case marking for predictive processing in second language Japanese. *Bilingualism: Language and Cognition*, *19*(1), 19–35. https://doi.org/10.1017/S1366728914000881

Montesano, L., Valentini, A., & Cornoldi, C. (2020). *LSC-SUA prove di lettura, comprensione del testo, scrittura e calcolo. Batteria per la valutazione dei DSA e altri disturbi in studenti universitari e adulti.*

Con 4 Fascicolo: Protocollo per l'esaminatore. Erickson. Retrieved October 26, 2023, from https://www.ibs.it/lsc-sua-prove-di-lettura-libro-vari/e/9788859020509

Moser, H., & Sedlaczek, R. (2020). *Das große Wörterbuch der Tiroler Dialekte*. Haymon Verlag.

Münster, K., & Knoeferle, P. (2018). Extending Situated Language Comprehension (Accounts) with Speaker and Comprehender Characteristics: Toward Socially Situated Interpretation. *Frontiers in Psychology*, *8*, 2267. https://doi.org/10.3389/fpsyg.2017.02267

Neely, J. H., & Kahan, T. A. (2001). Is semantic activation automatic? A critical re-evaluation. In *The nature of remembering: Essays in honor of Robert G. Crowder* (pp. 69–93). American Psychological Association. https://doi.org/10.1037/10394-005

Nieuwland, M. S., Politzer-Ahles, S., Heyselaar, E., Segaert, K., Darley, E., Kazanina, N., Von Grebmer Zu Wolfsthurn, S., Bartolozzi, F., Kogan, V., Ito, A., Mézière, D., Barr, D. J., Rousselet, G. A., Ferguson, H. J., Busch-Moreno, S., Fu, X., Tuomainen, J., Kulakova, E., Husband, E. M., ... Huettig, F. (2018). Large-scale replication study reveals a limit on probabilistic prediction in language comprehension. *eLife*, *7*, e33468. https://doi.org/10.7554/eLife.33468

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, *52*(3), 189–234. https://doi.org/10.1016/0010-0277(94)90043-4

Norton, E. S., & Wolf, M. (2012). Rapid Automatized Naming (RAN) and Reading Fluency: Implications for Understanding and Treatment of Reading Disabilities. *Annual Review of Psychology*, *63*(1), 427–452. https://doi.org/10.1146/annurev-psych-120710-100431

Odlin, T. (1989, June). *Language Transfer: Cross-Linguistic Influence in Language Learning* (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9781139524537

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, *31*(6), 785–806. https://doi.org/10.1016/0749-596X(92)90039-Z

Otten, M., Nieuwland, M. S., & Van Berkum, J. J. (2007). Great expectations: Specific lexical anticipation influences the processing of spoken language. *BMC Neuroscience*, *8*(1), 89. https://doi.org/10.1186/1471-2202-8-89

Otten, M., & Van Berkum, J. J. A. (2008). Discourse-Based Word Anticipation During Language Processing: Prediction or Priming? *Discourse Processes*, *45*(6), 464–496. https://doi.org/10.1080/01638530802356463

Otten, M., & Van Berkum, J. J. A. (2009). Does working memory capacity affect the ability to predict upcoming words in discourse? *Brain Research*, *1291*, 92–101. https://doi.org/10.1016/j.brainres.2009.07.042

Pacheco, A., Reis, A., Araújo, S., Inácio, F., Petersson, K. M., & Faísca, L. (2014). Dyslexia heterogeneity: Cognitive profiling of Portuguese children with dyslexia. *Reading and Writing*, *27*(9), 1529–1545. https://doi.org/10.1007/s11145-014-9504-5

Payesteh, B., & Pham, G. T. (2022). Is There a Cognate Effect in Bilingual Children With Developmental Language Disorder? *Language, Speech, and Hearing Services in Schools*, *53*(1), 213–221. https://doi.org/10.1044/2021_LSHSS-21-00078

Pennington, B. F. (2006). From single to multiple deficit models of developmental disorders. *Cognition*, *101*(2), 385–413. https://doi.org/10.1016/j.cognition.2006.04.008

Pennington, B. F., Santerre-Lemmon, L., Rosenberg, J., MacDonald, B., Boada, R., Friend, A., Leopold, D. R., Samuelsson, S., Byrne, B., Willcutt, E. G., & Olson, R. K. (2012). Individual prediction of dyslexia by single versus multiple deficit models. *Journal of Abnormal Psychology*, *121*(1), 212–224. https://doi.org/10.1037/a0025823

Peters, R. E., Grüter, T., & Borovsky, A. (2015). Anticipatory and Locally Coherent Lexical Activation Varies as a Function of Language Proficiency. *Cognitive Science*.

Pickering, M. J., & Gambi, C. (2018). Predicting while comprehending language: A theory and review. *Psychological Bulletin*, *144*(10), 1002–1044. https://doi.org/10.1037/bul0000158

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(4), 329–347. https://doi.org/10.1017/S0140525X12001495

Pickering, M. J., & Gompel, R. P. G. V. (2006). Syntactic Parsing. In *Handbook of Psycholinguistics* (pp. 455–503). Elsevier. https://doi.org/10.1016/B978-012369374-7/50013-4

Pisoni, D., & Levi, S. (2005). Some observations on representations and representational specificity in speech perception and spoken word recognition. Retrieved December 21, 2023, from https://www.semanticscholar.org/paper/Some-observations-on-representations-and-in-speech-Pisoni-Levi/c227d1ba395c630a36c6d76710c64f637ca8571c

Pisoni, D. B., & Levi, S. V. (2012, September). Representations and representational specificity in speech perception and spoken word recognition. In M. G. Gaskell (Ed.), *The Oxford Handbook of Psycholinguistics* (1st ed., pp. 3–18). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198568971.013.0001

Pulvermüller, F., & Grisoni, L. (2020). Semantic Prediction in Brain and Mind. *Trends in Cognitive Sciences*, *24*(10), 781–784. https://doi.org/10.1016/j.tics.2020.07.002

Ramus, F., & Szenkovits, G. (2008). What Phonological Deficit? *Quarterly Journal of Experimental Psychology*, *61*(1), 129–141. https://doi.org/10.1080/17470210701508822

Raney, G. E., Campbell, S. J., & Bovee, J. C. (2014). Using Eye Movements to Evaluate the Cognitive Processes Involved in Text Comprehension. *Journal of Visualized Experiments : JoVE*, (83), 50780. https://doi.org/10.3791/50780

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372–422. https://doi.org/10.1037/0033-2909.124.3.372

Rayner, K., & Well, A. D. (1996). Effects of contextual constraint on eye movements in reading: A further examination. *Psychonomic Bulletin & Review*, *3*(4), 504–509. https://doi.org/10.3758/BF03214555

Reis, A., Araújo, S., Morais, I. S., & Faísca, L. (2020). Reading and reading-related skills in adults with dyslexia from different orthographic systems: A review and meta-analysis. *Annals of Dyslexia*, *70*(3), 339–368. https://doi.org/10.1007/s11881-020-00205-x

Reiter, A., Tucha, O., & Lange, K. W. (2005). Executive functions in children with dyslexia. *Dyslexia*, *11*(2), 116–131. https://doi.org/10.1002/dys.289

Reuter, T., Borovsky, A., & Lew-Williams, C. (2019). Predict and redirect: Prediction errors support children's word learning. *Developmental Psychology*, *55*(8), 1656–1665. https://doi.org/10.1037/dev0000754

Robertson, E. K., & Joanisse, M. F. (2010). Spoken sentence comprehension in children with dyslexia and language impairment: The roles of syntax and working memory. *Applied Psycholinguistics*, *31*(1), 141–165. https://doi.org/10.1017/S0142716409990208

Rommers, J., Meyer, A. S., & Huettig, F. (2015). Verbal and nonverbal predictors of language-mediated anticipatory eye movements. *Attention, Perception, & Psychophysics*, *77*(3), 720–730. https://doi.org/10.3758/s13414-015-0873-x

Ryskin, R., Levy, R. P., & Fedorenko, E. (2020). Do domain-general executive resources play a role in linguistic prediction? Re-evaluation of the evidence and a path forward. *Neuropsychologia*, *136*, 107258. https://doi.org/10.1016/j.neuropsychologia.2019.107258

Saslow, M. G. (1967). Latency for Saccadic Eye Movement. *JOSA*, *57*(8), 1030–1033. https://doi.org/10.1364/JOSA.57.001030

Scarborough, D. L., Gerard, L., & Cortese, C. (1984). Independence of lexical access in bilingual word recognition. *Journal of Verbal Learning and Verbal Behavior*, *23*(1), 84–99. https://doi.org/10.1016/S0022-5371(84)90519-X

Schiff, R., & Raveh, M. (2007). Deficient morphological processing in adults with developmental dyslexia: Another barrier to efficient word recognition? *Dyslexia*, *13*(2), 110–129. https://doi.org/10.1002/dys.322

Schlenter, J. (2023). Prediction in bilingual sentence processing: How prediction differs in a later learned language from a first language. *Bilingualism: Language and Cognition*, *26*(2), 253–267. https://doi.org/10.1017/S1366728922000736

Schomers, M. R., Kirilina, E., Weigand, A., Bajbouj, M., & Pulvermüller, F. (2015). Causal Influence of Articulatory Motor Cortex on Comprehending Single Spoken Words: TMS Evidence. *Cerebral Cortex*, *25*(10), 3894–3902. https://doi.org/10.1093/cercor/bhu274

Schwartz, A. I., & Kroll, J. F. (2006). Bilingual lexical activation in sentence context. *Journal of Memory and Language*, *55*(2), 197–212. https://doi.org/10.1016/j.jml.2006.03.004

Schwartz, A. I., Kroll, J. F., & Diaz, M. (2007). Reading words in Spanish and English: Mapping orthography to phonology in two languages. *Language and Cognitive Processes*, *22*(1), 106–129. https://doi.org/10.1080/01690960500463920

Schwarz, J., Lizarazu, M., Lallier, M., & Klimovich-Gray, A. (2024). Phonological deficits in dyslexia impede lexical processing of spoken words: Linking behavioural and MEG data. *Cortex*, *171*, 204–222. https://doi.org/10.1016/j.cortex.2023.10.003

Scott, S. K., & Johnsrude, I. S. (2003). The neuroanatomical and functional organization of speech perception. *Trends in Neurosciences*, *26*(2), 100–107. https://doi.org/10.1016/S0166-2236(02)00037-1

Scott, S. K., McGettigan, C., & Eisner, F. (2009). A little more conversation, a little less action — candidate roles for the motor cortex in speech perception. *Nature Reviews Neuroscience*, *10*(4), 295–302. https://doi.org/10.1038/nrn2603

Segalowitz, N., & Hulstijn, J. (2009, February). Automaticity in Bilingualism and Second Language Learning. In J. F. Kroll & A. M. B. D. Groot (Eds.), *Handbook of Bilingualism* (pp. 371–388). Oxford University

PressNew York, NY. https://doi.org/10.1093/oso/9780195151770.003.0021

Setter, J., & Sebina, B. (2017, November). English lexical stress, prominence and rhythm. In O. Kang, R. I. Thomson, & J. M. Murphy (Eds.), *The Routledge Handbook of Contemporary English Pronunciation* (1st ed., pp. 137–153). Routledge. https://doi.org/10.4324/9781315145006-9

Simmons, E. S., & Magnuson, J. S. (2018). Word length, proportion of overlap, and phonological competition in spoken word recognition. *Proceedings of the 40th Annual Conference of the Cognitive Science Society*, 1062–1067.

Smith-Spark, J., Fisk, J., Fawcett, A., & Nicolson, R. (2003). Investigating the central executive in adult dyslexics: Evidence from phonological and visuospatial working memory performance. *European Journal of Cognitive Psychology*, *15*(4), 567–587. https://doi.org/10.1080/09541440340000024

Snowling, M. J. (2001). From language to reading and dyslexia. *Dyslexia (Chichester, England)*, *7*(1), 37–46. https://doi.org/10.1002/dys.185

Snowling, M. J. (2000). *Dyslexia, 2nd ed.* Blackwell Publishing.

Snowling, M. J., Hulme, C., & Nation, K. (2020). Defining and understanding dyslexia: Past, present and future. *Oxford Review of Education*, *46*(4), 501–513. https://doi.org/10.1080/03054985.2020.1765756

Spivey, M. (2002). Eye movements and spoken language comprehension: Effects of visual context on syntactic ambiguity resolution. *Cognitive Psychology*, *45*(4), 447–481. https://doi.org/10.1016/S0010-0285(02)00503-0

SR Research Ltd. (2005). EyeLink 1000. [Apparatus and software]. https://www.sr-research.com/

SR Research Ltd. (2018). Experiment Builder [Computer software]. https://www.sr-research.com/experiment-builder/

Stanovich, K. E. (1986). Matthew Effects in Reading: Some Consequences of Individual Differences in the Acquisition of Literacy. *Reading Research Quarterly*, *21*(4), 360–407. Retrieved August 22, 2025, from https://www.jstor.org/stable/747612

Stanovich, K. E., & West, R. F. (1989). Exposure to print and orthographic processing. *Reading Research Quarterly*, *24*(4), 402–433. https://doi.org/10.2307/747605

Staub, A., & Clifton, C. (2006). Syntactic prediction in language comprehension: Evidence from either...or. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(2), 425–436. https://doi.org/10.1037/0278-7393.32.2.425

Staub, A., Grant, M., Astheimer, L., & Cohen, A. (2015). The influence of cloze probability and item constraint on cloze task response time. *Journal of Memory and Language*, *82*, 1–17. https://doi.org/10.1016/j.jml.2015.02.004

Stella, M., & Engelhardt, P. E. (2019). Syntactic ambiguity resolution in dyslexia: An examination of cognitive factors underlying eye movement differences and comprehension failures. *Dyslexia*, *25*(2), 115–141. https://doi.org/10.1002/dys.1613

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643–662. https://doi.org/10.1037/h0054651

Tagliani, M., Cruz, L., Redolfi, M., Shirokorad, N., Canzi, M., Melloni, C., & Vender, M. (2025). L2 learners take more time to catch the rhyme: An eye-tracking study on predictive processing. *International Journal of Bilingualism*, 13670069251365147. https://doi.org/10.1177/13670069251365147

Tanenhaus, M. K., & Trueswell, J. C. (1995). Sentence Comprehension. In *Speech, Language, and Communication* (pp. 217–262). Elsevier. https://doi.org/10.1016/B978-012497770-9.50009-1

Taylor, W. L. (1953). "Cloze procedure": A new tool for measuring readability. *Journalism quarterly*, *30*(4), 415–433.

Titone, D., Libben, M., Mercier, J., Whitford, V., & Pivneva, I. (2011). Bilingual lexical access during L1 sentence reading: The effects of L2 knowledge, semantic constraint, and L1–L2 intermixing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(6), 1412–1431. https://doi.org/10.1037/a0024492

Trueswell, J., Tanenhaus, M., & Garnsey, S. (1994). Semantic Influences On Parsing: Use of Thematic Role Information in Syntactic Ambiguity Resolution. *Journal of Memory and Language*, *33*(3), 285–318. https://doi.org/10.1006/jmla.1994.1014

Tullio De Mauro. (2016). Dizionario italiano De Mauro - Vocabolario online della lingua italiana. Retrieved October 1, 2023, from https://dizionario.internazionale.it/www.internazionale.it

Tyler, L. K. (1984). The structure of the initial cohort: Evidence from gating. *Perception & Psychophysics*, *36*(5), 417–427. https://doi.org/10.3758/BF03207496

Van Assche, E., Drieghe, D., Duyck, W., Welvaert, M., & Hartsuiker, R. J. (2011). The influence of semantic constraints on bilingual word recognition during sentence reading. *Journal of Memory and Language*, *64*(1), 88–107. https://doi.org/10.1016/j.jml.2010.08.006

Van Bergen, G., & Flecken, M. (2017). Putting things in new places: Linguistic experience modulates the predictive power of placement verb semantics. *Journal of Memory and Language*, *92*, 26–42. https://doi.org/10.1016/j.jml.2016.05.003

Van Berkum, J. J. A., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating Upcoming Words in Discourse: Evidence From ERPs and Reading Times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(3), 443–467. https://doi.org/10.1037/0278-7393.31.3.443

Van Berkum, J. J. A., De Goede, D., Van Alphen, P. M., Mulder, E. R., & Kerstholt, J. H. (2013). How robust is the language architecture? The case of mood. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00505

Van Berkum, J. J. A., Van Den Brink, D., Tesink, C. M. J. Y., Kos, M., & Hagoort, P. (2008). The Neural Integration of Speaker and Message. *Journal of Cognitive Neuroscience*, *20*(4), 580–591. https://doi.org/10.1162/jocn.2008.20054

Van Hell, J. G., & De Groot, A. M. (2008). Sentence context modulates visual word recognition and translation in bilinguals. *Acta Psychologica*, *128*(3), 431–451. https://doi.org/10.1016/j.actpsy.2008.03.010

Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology*, *83*(2), 176–190. https://doi.org/10.1016/j.ijpsycho.2011.09.015

Van Rij, J., Wieling, M., & Baayen, R. H. (2015, February). Itsadug: Interpreting Time Series and Autocorrelated Data Using GAMMs [Institution: Comprehensive R Archive Network Pages: 2.4.1]. https://doi.org/10.32614/CRAN.package.itsadug

Vellutino, F. R., Fletcher, J. M., Snowling, M. J., & Scanlon, D. M. (2004). Specific reading disability (dyslexia): What have we learned in the past four decades? *Journal of Child Psychology and Psychiatry*, *45*(1), 2–40. https://doi.org/10.1046/j.0021-9630.2003.00305.x

Vender, M., & Delfitto, D. (2010). Towards a pragmatics of negation: The interpretation of negative sentences in developmental dyslexia. *Generative Grammar in Geneva*, *6*, 1–27.

Vender, M., & Delfitto, D. (2024). Bridging the Gap in Adult Dyslexia Research: Assessing the Efficacy of a Linguistic Intervention on Literacy Skills. *Annals of Dyslexia*. https://doi.org/10.1007/s11881-024-00314-x

Vender, M., & Nardon, A. (2023). Un'analisi del vocabolario ortografico e fonologico di apprendenti italiani di inglese: Il ruolo delle parole imparentate. *Studi di Glottodidattica*, (2), 99–115.

Vettori, C., Colombo, S., & Abel, A. (2021). Multilingualism in South Tyrol: Between old fears and new challenges. *Zeitschrift für Interkulturellen Fremdsprachenunterricht*, *26*(2), 223–249.

Vettori, C., Wisniewski, K., & Abel, A. (2012, August). Becoming bilingual in a multilingual context: A snapshot view of L2 competences in South Tyrol. In K. Braunmüller & C. Gabriel (Eds.), *Hamburg Studies on Multilingualism* (pp. 437–456, Vol. 13). John Benjamins Publishing Company. https://doi.org/10.1075/hsm.13.29vet

Vietti, A., Alber, B., & Vogt, B. (2018). Initial laryngeal neutralisation in Tyrolean. *Phonology*, *35*(1), 79–114. https://doi.org/10.1017/S0952675717000380

Vukovic, R. K., Wilson, A. M., & Nash, K. K. (2004). Naming Speed Deficits in Adults with Reading Disabilities: A Test of the Double-Deficit Hypothesis. *Journal of Learning Disabilities*, *37*(5), 440–450. https://doi.org/10.1177/00222194040370050601

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., & Pearson, N. A. (2013). Comprehensive Test of Phonological Processing–Second Edition (CTOPP-2) [Institution: American Psychological Association]. https://doi.org/10.1037/t52630-000

Wagner, R. K., Zirps, F. A., Edwards, A. A., Wood, S. G., Joyner, R. E., Becker, B. J., Liu, G., & Beal, B. (2020). The Prevalence of Dyslexia: A New Approach to Its Estimation. *Journal of Learning Disabilities*, *53*(5), 354–365. https://doi.org/10.1177/0022219420920377

Walenski, M., Europa, E., Caplan, D., & Thompson, C. K. (2019). Neural networks for sentence comprehension and production: An ALE-based meta-analysis of neuroimaging studies. *Human Brain Mapping*, *40*(8), 2275–2304. https://doi.org/10.1002/hbm.24523

Watkins, K. E., Strafella, A. P., & Paus, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia*, *41*(8), 989–994. https://doi.org/10.1016/S0028-3932(02)00316-0

Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language*, *50*(1), 1–25. https://doi.org/10.1016/S0749-596X(03)00105-0

Wicha, N. Y. Y., Moreno, E. M., & Kutas, M. (2004). Anticipating Words and Their Gender: An Event-related Brain Potential Study of Semantic Integration, Gender Expectancy, and Gender Agreement in Spanish Sentence Reading. *Journal of Cognitive Neuroscience*, *16*(7), 1272–1288. https://doi.org/10.1162/0898929041920487

Wieling, M., Margaretha, E., & Nerbonne, J. (2012). Inducing a measure of phonetic similarity from pronunciation variation. *Journal of*

Phonetics, *40*(2), 307–314. https://doi.org/10.1016/j.wocn.2011.12.004

Wiseheart, R., Altmann, L. J. P., Park, H., & Lombardino, L. J. (2009). Sentence comprehension in young adults with developmental dyslexia. *Annals of Dyslexia*, *59*(2), 151–167. https://doi.org/10.1007/s11881-009-0028-7

Wlotko, E. W., & Federmeier, K. D. (2012). Age-related changes in the impact of contextual strength on multiple aspects of sentence comprehension. *Psychophysiology*, *49*(6), 770–785. https://doi.org/10.1111/j.1469-8986.2012.01366.x

Wolf, M., & Bowers, P. G. (1999). The double-deficit hypothesis for the developmental dyslexias. *Journal of Educational Psychology*, *91*(3), 415–438. https://doi.org/10.1037/0022-0663.91.3.415

Wood, S. N. (2017, May). *Generalized Additive Models: An Introduction with R, Second Edition* (2nd ed.). Chapman; Hall/CRC. https://doi.org/10.1201/9781315370279

Yang, L., Li, C., Li, X., Zhai, M., An, Q., Zhang, Y., Zhao, J., & Weng, X. (2022). Prevalence of Developmental Dyslexia in Primary School Children: A Systematic Review and Meta-Analysis. *Brain Sciences*, *12*(2), 240. https://doi.org/10.3390/brainsci12020240

Yee, E., & Sedivy, J. C. (2006). "Eye movements to pictures reveal transient semantic activation during spoken word recognition": Correction to Yee and Sedivy (2006). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(2), ii–ii. https://doi.org/10.1037/0278-7393.32.2.ii

Yin, H., & Pickering, M. J. (2026). Predicting words across languages depends on language context: Evidence from visual world eye-tracking. *Journal of Memory and Language*, *146*, 104701. https://doi.org/10.1016/j.jml.2025.104701

Zachou, A., Partesana, E., Tenca, E., & Guasti, M. T. (2013). Production and comprehension of direct object clitics and definite articles by Italian children with developmental dyslexia. *Advances in language acquisition*, 464–471.

Ziegler, J. C., Bertrand, D., Tóth, D., Csépe, V., Reis, A., Faísca, L., Saine, N., Lyytinen, H., Vaessen, A., & Blomert, L. (2010). Orthographic Depth and Its Impact on Universal Predictors of Reading: A Cross-Language Investigation. *Psychological Science*, *21*(4), 551–559. https://doi.org/10.1177/0956797610363406

Ziegler, J. C., & Goswami, U. (2005). Reading Acquisition, Developmental Dyslexia, and Skilled Reading Across Languages: A Psycholinguistic Grain Size Theory. *Psychological Bulletin*, *131*(1), 3–29. https://doi.org/10.1037/0033-2909.131.1.3

Appendices

Appendix A

Stimuli for Experiment 1a

A.1 Experimental Stimuli - Experiment 1a

Table A.1: Complete list of experimental stimuli with English translations. Target words are **bold** in both Italian and English sentences. Filler items (IDs ending with "b") were paired with the same target, phonological competitor, control, and distractor images as their corresponding experimental sentences, but their item columns are omitted for brevity.

#	Italian Sentence	English Translation	Phon. Comp.	Control	Distr. 1	Distr. 2	Distr. 3
1	Nel film d'azione l'eroe sfugge e si lancia dall'aereo con un **paracadute**.	In the action movie the hero escapes and jumps from the plane with a **parachute**.	pavone (peacock)	dentifricio (toothpaste)	ragno (spider)	criceto (hamster)	trattore (tractor)
1b	Vado a trovare i miei nonni in campagna e vedo un **trattore**.	I go to visit my grandparents in the countryside and see a **tractor**.	*Filler sentence*				
2	L'elettricista sostituisce il lampadario e sale sulla **scala**.	The electrician replaces the chandelier and climbs the **ladder**.	scatola (box)	pianta (plant)	lettera (letter)	bottiglia (bottle)	palla (ball)
2b	Laura è molto disordinata e sul divano lascia una **forchetta**.	Laura is very messy and leaves a **fork** on the sofa.	*Filler sentence*				

Table continues on the next page

Table A.1: Experimental stimuli — continued

#	Italian Sentence	English Translation	Phon. Comp.	Control	Distr. 1	Distr. 2	Distr. 3
3	Marta prepara la frittata e mette un filo d'olio nella **padella**.	Marta prepares the omelette and puts a drop of oil in the **pan**.	paletta (dustpan)	foca (seal)	ciliegia (cherry)	calcolatrice (calculator)	torcia (flashlight)
3b	Entro in casa e vedo mio padre con in mano una **sigaretta**.	I enter the house and see my father with a **cigarette** in his hand.			*Filler sentence*		
4	Vado in campagna e pedalo per chilometri sulla mia **bicicletta**.	I go to the countryside and pedal for kilometers on my **bicycle**.	bilancia (scale)	matita (pencil)	mela (apple)	chiave (key)	giacca (jacket)
4b	Prima di uscire Anna lava i denti e prende la **sciarpa**.	Before going out Anna brushes her teeth and takes the **scarf**.			*Filler sentence*		
5	Dopo la pioggia con il sole alziamo gli occhi e vediamo l'**arcobaleno**.	After the rain with the sun we raise our eyes and see the **rainbow**.	armadio (wardrobe)	insalata (salad)	imbuto (funnel)	orso (bear)	ostrica (oyster)
5b	Apro una rivista e vedo l'immagine di un **elefante**.	I open a magazine and see the image of an **elephant**.			*Filler sentence*		

Table continues on the next page

Table A.1: Experimental stimuli — continued

#	Italian Sentence	English Translation	Phon. Comp.	Control	Distr. 1	Distr. 2	Distr. 3
6	Il bambino ha tanta tosse e deve prendere lo **sciroppo**	The child has a bad cough and needs to take some cough **syrup**.	sciatore (skier)	gnomo (gnome)	spazzolino (toothbrush)	zaino (backpack)	smalto (nail polish)
6b	Martina è molto arrabbiata oggi e lancia lo **zaino**.	Martina is very angry today and throws the **backpack**.			*Filler sentence*		
7	La piccola bambina ride felice e si dondola sull'**altalena**.	The little girl laughs happily and swings on the **swing**.	albicocca (apricot)	elicottero (helicopter)	ombrellone (beach umbrella)	ippopotamo (hippo)	orecchino (earring)
7b	Prendo il mio telefono e cerco l'immagine di un **asino**.	I take my phone and look for the picture of a **donkey**.			*Filler sentence*		
8	Per le vacanze vado nel deserto e cavalco un **cammello**.	For vacation I go to the desert and ride a **camel**.	campanile (bell tower)	cesto (basket)	astuccio (pencil case)	palloncino (balloon)	dentifricio (toothpaste)
8b	Federico pulisce la casa e nell'armadio trova un **telecomando**.	Federico cleans the house and finds a **remote control** in the wardrobe.			*Filler sentence*		
9	I bambini giocano in spiaggia e costruiscono **castelli**	The children play on the beach and build **sandcastles**.	castori (beavers)	giornali (newspapers)	occhiali (glasses)	fiori (flowers)	piatti (plates)

Table continues on the next page

Table A.1: Experimental stimuli — continued

#	Italian Sentence	English Translation	Phon. Comp.	Control	Distr. 1	Distr. 2	Distr. 3
9b	Tutto il giorno mia nonna gira per casa e cerca i suoi **occhiali**.	All day my grandmother walks around the house and looks for her **glasses**.			*Filler sentence*		
10	Prendo appunti e macchio il mio abito con l'inchiostro della **penna**.	I take notes and stain my dress with the **pen** ink.	pentola (pot)	maschera (mask)	zucca (pumpkin)	torta (cake)	lampada (lamp)
10b	Vado nel centro commerciale sotto casa e compro una **torta**.	I go to the shopping center under my house and buy a **cake**.			*Filler sentence*		
11	Laura si prepara per il Natale e attacca le calze al **camino**.	Laura prepares for Christmas and hangs the stockings on the **fireplace**.	canestro (basket hoop)	pennarello (marker)	dinosauro (dinosaur)	tamburo (drum)	semaforo (traffic light)
11b	Pietro frequenta la lezione di arte e dipinge un **tamburo**.	Pietro attends art class and paints a **drum**.			*Filler sentence*		
12	Finisco la cena e lascio una generosa mancia al **cameriere**.	I finish dinner and leave a generous tip to the **waiter**.	calendario (calendar)	gelato (ice cream)	bottone (button)	microfono (microphone)	pomodoro (tomato)
12b	La ragazza perde l'equilibrio e lascia cadere il **microfono**.	The girl loses her balance and drops the **microphone**.			*Filler sentence*		

Table continues on the next page

Table A.1: Experimental stimuli — continued

#	Italian Sentence	English Translation	Phon. Comp.	Control	Distr. 1	Distr. 2	Distr. 3
13	Conosco un ottimo caseificio dove compro sempre il **formaggio**.	I know an excellent cheese factory where I always buy **cheese**.	forno (oven)	secchio (bucket)	violino (violin)	materasso (mattress)	telescopio (telescope)
13b	Ho molte cose inutili in garage e in soffitta ho un **violino**.	I have many useless things in the garage and in the attic I have a **violin**.			*Filler sentence*		
14	Veronica si trucca gli occhi e sul collo si spruzza il **profumo**.	Veronica puts makeup on her eyes and sprays **perfume** on her neck.	prosciutto (ham)	leone (lion)	vaso (vase)	gelato (ice cream)	cuscino (pillow)
14b	Dopo il lavoro passo al supermercato e compro il **riso**.	After work I stop by the supermarket and buy **rice**.			*Filler sentence*		
15	Un uccello raccoglie la paglia e costruisce il nido sull'**albero**.	A bird collects straw and builds the nest on the **tree**.	album (fotografico) (photo album)	biscotto (cookie)	ombrello (umbrella)	topo (mouse)	pirato (pirate)
15b	La mamma lava i piatti e la bambina disegna un **topo**.	Mom washes the dishes and the little girl draws a **mouse**.			*Filler sentence*		

Table continues on the next page

Table A.1: Experimental stimuli — continued

#	Italian Sentence	English Translation	Phon. Comp.	Control	Distr. 1	Distr. 2	Distr. 3
16	Alice va a lezione e scrive i suoi appunti in un **quaderno**.	Alice goes to class and writes her notes in a **notebook**.	quadrato (square)	cucchiaio (spoon)	fucile (rifle)	drago (dragon)	mantello (cloak)
16b	Carlo scrive una lettera a Babbo Natale e chiede un **drago**.	Carlo writes a letter to Santa Claus and asks for a **dragon**.			*Filler sentence*		
17	La sera vado a correre e tengo il telefonino in **tasca**.	In the evening I go running and keep my phone in my **pocket**.	tastiera (keyboard)	moneta (coin)	collana (necklace)	sedia (chair)	gabbia (cage)
17b	Massimo va al mercatino dell'usato e cerca una **collana**.	Massimo goes to the flea market and looks for a **necklace**.			*Filler sentence*		
18	La notte sento un fastidioso ronzio e vengo punta da una **zanzara**.	At night I hear an annoying buzzing and get stung by a **mosquito**.	zanna (tusk)	conchiglia (seashell)	banana (banana)	fionda (slingshot)	lumaca (snail)
18b	Ogni notte faccio un incubo terribile e sogno una **lumaca**.	Every night I have a terrible nightmare and dream of a **snail**.			*Filler sentence*		
19	A Capodanno i ragazzi si divertono a far scoppiare i **petardi**.	At New Year's the boys have fun setting off **firecrackers**.	peperoni (bell peppers)	delfini (dolphins)	cerotti (band-aids)	sandali (sandals)	tulipani (tulips)

Table continues on the next page

Table A.1: Experimental stimuli — continued

#	Italian Sentence	English Translation	Phon. Comp.	Control	Distr. 1	Distr. 2	Distr. 3
19b	La donna chiama il marito e gli chiede di comprare i **cerotti**.	The woman calls the husband and asks him to buy **band-aids**.			*Filler sentence*		
20	Pietro apre una bottiglia di spumante e fa saltare via il **tappo**.	Pietro opens a bottle of sparkling wine and pops the **cork**.	tappeto (rug)	palloncino (balloon)	pinguino (penguin)	pompiere (firefighter)	rastrello (rake)
20b	Osservo il cielo e vedo una nuvola simile a un **pinguino**.	I observe the sky and see a cloud similar to a **penguin**.			*Filler sentence*		
21	Mescolo il minestrone e copro la pentola con il **coperchio**.	I stir the minestrone and cover the pot with the **lid**.	coniglio (rabbit)	fungo (mushroom)	motorino (scooter)	pipistrello (bat)	tostapane (toaster)
21b	Una mattina Guido accende la radio e parlano di un **tostapane**.	One morning Guido turns on the radio and they talk about a **toaster**.			*Filler sentence*		
Practice trials.							
P1	Finalmente Sara supera l'esame di guida e compra una **macchina**.	Sara finally passes the driving test and buys a **car**.			chiesa (church)	lingua (tongue)	porta (door)

Table continues on the next page

Table A.1: Experimental stimuli — continued

#	Italian Sentence	English Translation	Phon. Comp.	Control	Distr. 1	Distr. 2	Distr. 3
P2	L'anziano signore passeggia nel parco e si siede sulla **panchina**.	The elderly gentleman walks in the park and sits on the **bench**.	pantofola (slipper)		bambola (doll)	spada (sword)	stella (star)
P3	Il bagnino osserva il mare e studia la forma dell'**onda**.	The lifeguard watches the sea and studies the shape of the **wave**.		ago (needle)	erba (grass)	uovo (egg)	osso (bone)
P4	La sera accendo la tivù e c'è la pubblicità di un **detersivo**.	In the evening I turn on the TV and there is an advertisement for a **detergent**.		detersivo (detergent)	orologio (watch)	vestito (dress)	bollitore (kettle)

End of table

A.2 Experimental Sentence Transcriptions - Experiment 1a

Table A.2: Phonetic transcriptions of experimental stimuli.

#	**Phonetic Transcription**
1	[nelˈfilmdat͡sˈt͡sːjone leˈrɔːe ˈsfud͡ʒd͡ʒːe esiˈlant͡ʃadallaˈɛːreo konumparakaˈduːte]
2	[lelettriˈt͡ʃiːsta sostituˈiːʃeillampaˈdaːrjo eˈsalesullaˈskaːla]
3	[ˈmartapreˈpaːralafritˈtaːta eˈmetteunˈfilodˈoːljo ˈnellapaˈdella]
4	[ˈvadoiŋkkamˈpaɲɲːa epedˈdaːloperkiˈlɔːmetri ˈsullaˈmiːabit͡ʃiˈkletta]
5	[ˈdoːpolaˈpjɔd͡ʒd͡ʒːakonilˈsole alˈt͡sjamoʎˈʎːokkːi eveˈdjaːmolarkobaˈlɛːno]
6	[ilbamˈbiːno atˈtantaˈtosse edˈdɛːveˈprɛndereloʃʃːiˈrɔppo]
7	[laˈpikkːolabamˈbiːna ˈriːdefeˈliːt͡ʃe esiˈdondolasullːaltaˈleːna]
8	[perlevaˈkant͡se ˈvadoneldeˈzɛrto ekaˈvalkouŋkamˈmːɛllːo]
9	[ibamˈbiniˈd͡ʒoːkanoinˈspjad͡ʒd͡ʒːa ekostruˈiːskonokaˈstelli]
10	[ˈprɛndoapˈpunti eˈmakkjo ilˈmiːoˈaːbito konlinˈkjɔstro dellaˈpenna]
11	[ˈlawra sipreˈpaːraperilnaˈtaːle eaˈttakkaleˈkalt͡st͡sːe alkaˈmːino]
12	[fiˈniskolat͡ʃˈt͡ʃeːna eˈlaʃʃːounad͡ʒːeneˈroːzaˈmant͡ʃa alkameˈrjɛːre]
13	[koˈnoskounˈottimokazejˈfiːt͡ʃo ˈdoːveˈkompro ˈsempre ilforˈmad͡ʒd͡ʒːo]
14	[veˈrɔːnikasiˈtrukkːaʎʎːiˈɔkkːi esulˈkɔllːosiˈsprut͡st͡sːailproˈfumo]
15	[unut͡ʃˈt͡ʃːɛllorakˈkːoʎʎːelaˈpaʎʎːa ekkːostruˈiʃʃːeilˈniːdosulˈlːalbero]
16	[aˈlit͡ʃevaːlet͡sˈt͡sːjoːne esˈkriːveiswɔjapˈpunti inuŋkwaˈdɛrno]
17	[laˈseːraˈvaːdoaˈkorrːere eˈtɛŋgoiltelefoˈniːnoinˈtaska]
18	[laˈnottːe ˈsentounfastiˈdjoːzoronˈd͡ziːo eˈvengoˈpunta daunad͡zd͡zːanˈd͡zara]
19	[akkapoˈdanno iraˈgat͡st͡sːisidiˈvertono afarskopˈpːjaːreipeˈtardi]
20	[ˈpjɛːtro apreunabotˈtiʎʎːadispuˈmante effːasalˈtaːreˈviailˈtappːo]
21	[ˈmɛːskoloilmineˈstroːne eˈkɔːprolaˈpɛntolakonilkoˈpɛrkjo]

End of table

Appendix B

Stimuli for Experiment 1b

B.1 Experimental Stimuli - Experiment 1b

Table B.1: Complete list of experimental stimuli for Experiment 1b with Tyrolean and English translations. All trials began with “Guarda la loro” (Look at their). Targets are **bold**. *Cognates* are in italics. English translations are shown in square brackets. Tyrolean orthography follows Moser and Sedlaczek (2020).

#	Italian Target	Tyrolean Target	Italian Competitors	Tyrolean Competitors
Prediction Congruent Condition				
1	**la mucca (F)** [cow]	**die Kuah (F)**	il palloncino (M) [balloon], il calzino (M) [sock], il frigorifero (M) [refrigerator]	der Luftballon (M), der Pfose/Sock (M), der Eiskoschtn (M)
2	***il pinguino* (M)** [penguin]	***der Pinguin* (M)**	la ragnatela (F) [spider web], *la sega* (F) [saw], la spazzola (F) [brush]	die Spinnewettn (F), *die Säge*/die Sog (F), die Birscht (F)
3	***il leone* (M)** [lion]	***der Löwe* (M)**	la scatola (F) [box], *la chitarra* (F) [guitar], la bottiglia (F) [bottle]	die Schochtl/Gschtåttl (F), *die Gitarre* (F), die Flosch (F)
4	**la candela (F)** [candle]	**die Kerz (F)**	il guanto (M) [glove], il fungo (M) [mushroom], il cucchiaio (M) [spoon]	der Hantschuach (M), der Pilz/Schwåmm (M), der Leffl (M)
5	***il pappagallo* (M)** [parrot]	***der Papagei* (M)**	la pipa (F) [pipe], *la banana* (F) [banana], la forchetta (F) [fork]	die Pfeife (F), *die Banane* (F), die Gobl (F)

Table continues on the next page

Table B.1: Experimental stimuli 1b — continued

#	Italian Target	Tyrolean Target	Italian Competitors	Tyrolean Competitors
6	**la pera (F)** [pear]	**die Birne (F)**	il martello (M) [hammer], il miele (M) [honey], il cappello (M) [hat]	der Hommr (M), der Henig/Honig (M), der Huat (M)
Prediction Incongruent Condition				
7	**il gatto (M)** [cat]	**die Kotz (F)**	la scarpa (F) [shoe], la mela (F) [apple], la pentola (F) [pot]	der Schuah (M), der Epfl (M), der Topf (M)
8	**il semaforo (M)** [traffic light]	**die Ompl (F)**	la scopa (F) [broom], la gonna (F) [skirt], la valigia (F) [suitcase]	der Besn (M), der Rock (M), der Koffer/Gufer/Kufer (M)
9	***il materasso* (M)** [mattress]	***die Matrotz* (F)**	la matita (F) [pencil], la cintura (F) [belt], *la tigre* (F) [tiger]	der Bleischtift (M), der Girtl/Gürtel (M), *der Tiger* (M)
10	***la palla* (F)** [ball]	***der Ball* (M)**	il fiore (M) [flower], il sapone (M) [soap], *il cioccolato* (M) [chocolate]	die Bloam (F), die Soaf (F), *die Schokolade/Tschuglatt* (F)
11	**la farfalla (F)** [butterfly]	**der Schmetterling/ Pfeifolter (M)**	il pomodoro (M) [tomato], il latte (M) [milk], il topo (M) [mouse]	die Tomaat (F), die Milch (F), die Maus (F)
12	***la stella* (F)** [star]	***der Stern* (M)**	il pipistrello (M) [bat], il violino (M) [violin], *il vaso* (M) [vase]	die Fledermaus (F), die Geige (F), *die Vase* (F)
Control Condition (No Prediction)				
13	***la lampada* (F)** [lamp]	***die Lomp* (F)**	la salsiccia (F) [sausage], la capra (F) [goat], *la corona* (F) [crown]	die Wurscht/Wurz (F), die Goaß (F), *die Krone* (F)

Table continues on the next page

Table B.1: Experimental stimuli 1b — continued

#	Italian Target	Tyrolean Target	Italian Competitors	Tyrolean Competitors
14	**il treno (M)** [train]	**der Zug (M)**	il formaggio (M) [cheese], il tavolo (M) [table], il cane (M) [dog]	der Kaas/Käse (M), der Tisch (M), der Hund (M)
15	***il kiwi* (M)** [kiwi]	***der Kiwi* (M)**	il maglione (M) [sweater], il criceto (M) [hamster], l'ombrello (M) [umbrella]	der Pullover/Pengger/Schwetter (M), der Homschter (M), der Schirm (M)
16	**la pala (F)** [shovel]	**die Schaufl (F)**	la conchiglia (F) [seashell], *la carota* (F) [carrot], la ghianda (F) [acorn]	die Muschl (F), *die Karotte* (F), die Oachl (F)
17	**la ciliegia (F)** [cherry]	**die Kerscht (F)**	la bambola (F) [doll], *la pizza* (F) [pizza], *la tazza* (F) [cup]	die Puppe (F), *die Pizza* (F), *die Tasse* (F)
18	**il casco (M)** [helmet]	**der Helm (M)**	il secchio (M) [bucket], il pettine (M) [comb], il gallo (M) [rooster]	der Kiebl (M), der Kompl (M), der Gigger/Giggl (M)
Practice Trials				
P1	**la collana (F)** [necklace]	**die Kette (F)**	*il camino* (M) [fireplace], *il vino* (M) [wine], l'albero (M) [tree]	*der Kem/Kamin* (M), *der Wein* (M), der Bam (M)
P2	***il tamburo* (M)** [drum]	***die Tromml* (F)**	la sedia (F) [chair], la chiave (F) [key], la patata (F) [potato]	der Schtual (M), der Schlissl (M), der Eardepfl (M)
P3	**la chiesa (F)** [church]	**die Kirch (F)**	*la giraffa* (F) [giraffe], la scala (F) [ladder], l'oca (F) [goose]	*die Giraffe* (F), die Loater (F), die Gonz (F)

End of table

Appendix C

Stimuli for Experiment 2

C.1 Experimental Stimuli — Experiment 2

Table C.1: Complete list of experimental stimuli for Experiment 2. Target words are **bold** in the sentence. Semantic competitors are annotated with subtype indicators in parentheses: CP+Sem.

#	Sentence	Phon. Comp.	Semantic Competitor	Unrelated
1	The computer pointer doesn't move although Tommy is moving the **mouse** quickly.	mouth	cable (CP+Sem)	fan
2	The tourists expected rain when the sun went behind the **clouds** but the weather got better later.	clown	eagle (CP-only)	brownie
3	Amelia got a driving license and will buy her own **car** tomorrow.	card	tyre (CP+Sem)	peppermill
4	Dillan got lost today so tomorrow he will use his **map** to find a way.	mannequin	globe (CP+Sem)	crocodile
5	*Item removed during stimulus cleaning.*			
6	At the train station the commuters are rushing to buy a **ticket** before the train leaves.	tissue box	QRcode (CP+Sem)	gorilla
7	The restaurant is always busy so leo will book a **table** for the whole family.	tape	drawers (CP-only)	puzzle piece

Table continues on the next page

Table C.1: Experimental stimuli — continued

#	Sentence	Phon. Comp.	Semantic Competitor	Unrelated
8	Kate didn't like coffee so she ordered a cup of **tea** with her cake.	teeth	lemon (CP-only)	kite
9	Meg will go to the park to walk her **dog** this afternoon.	dolphin	frisbee (CP+Sem)	binder
10	The mole was digging a small **hole** in the back yard.	hose	drill (CP-only)	celery
11	Hearing the noise outside the classroom the lecturer closed the **door** so that the students can concentrate better.	doughnut	keyhole (CP-only)	hairband
12	People saw the first spaceship that landed on the **moon** in the space museum.	mole	owl (CP-only)	shield
13	Mia got cold and has a runny **nose** and so does her brother.	notebook	ear (CP-only)	pickle
14	Rose couldn't eat noodles using chopsticks so used a **fork** instead.	folder	napkin (CP+Sem)	shirt
15	When its sunny Sarah always covers her head and normally wears a baseball **cap** outside.	cat	glove (CP+Sem)	trampoline
16	*Item removed during stimulus cleaning.*			

Table continues on the next page

Table C.1: Experimental stimuli — continued

#	Sentence	Phon. Comp.	Semantic Competitor	Unrelated
17	For Christmas the children are hanging bells on the **tree** and putting the stocking by the fire.	tricycle	stump (CP-only)	dice
18	To finish the cake scarlet spread the whipped **cream** smoothly over the top.	cross	cheese (CP+Sem)	nail
19	So she could rub out any mistakes in her notes Beth decided to write with a **pencil** for all her classes this term.	pengin	clipboard (CP+Sem)	button
20	Harry intends to propose to Emily and give her the **ring** on Saturday.	rink	hand (CP+Sem)	butter
21	As a life time vegetarian Olivia doesn't miss eating **meat** at all.	mirror	cow (CP+Sem)	hanger
22	The expensive wine is made from special type of **grape** that is grown only in the south of France.	grave	banana (CP-only)	slipper
23	On the island Jim went swimming and got a sun tan at the **beach** after just two hours.	beaker	lifesaver (CP+Sem)	mug
24	The child believed that Santa Claus would come into her house down the **chimney** at midnight.	chili	balcony (CP-only)	screw

Table continues on the next page

Table C.1: Experimental stimuli — continued

#	Sentence	Phon. Comp.	Semantic Competitor	Unrelated
25	The traveler went to the desert because he wanted to ride a **camel** and go exploring.	carrot	zebra (CP-only)	lamp
26	Kim found the room was too hot and humid so to get some fresh air she opened the **window** fully.	windmill	picture frame (CP+Sem)	razor
27	Instead of sending an email Ellie decided to write a **letter** to her sister.	lettuce	postbox (CP+Sem)	crab
28	In an emergency we cannot use a lift instead we need to use the **stairs** for our safety.	stapler	ladder (CP+Sem)	poker chips
29	Hannah brought a calendar and hung it on the **wall** next to her desk.	water	light switch (CP-only)	recorder
30	The camp leader taught children how to pitch a **tent** on their own.	tennis ball	house (CP-only)	bowl
31	All the employers here have saving accounts at the same **bank** for some reason.	bag	safe (CP+Sem)	amber
32	Lewis lost his memory because of the damage to his **brain** from his stroke.	bread	hard hat (CP+Sem)	coffee maker
33	Voilet left the dirty plates and cups in the **sink** all day.	syringe	towel (CP+Sem)	bridge

Table continues on the next page

Table C.1: Experimental stimuli — continued

#	Sentence	Phon. Comp.	Semantic Competitor	Unrelated
34	If you put water in the freezer it will become **ice** within a few hours.	iron	snowman (CP+Sem)	bolt
35	In the tennis lesson Lauran hit the ball with her **racket** very hard.	raspberry	sieve (CP-only)	curtain
36	Karen could not get the volleyball shot over the **net** as she was small.	neck	fence (CP-only)	telescope
37	The web had been spun by a large **spider** and it had already caught a lot of insects.	spice rack	ant (CP-only)	crayon
38	Sam didn't know the time because he forgot to wear his **watch** today.	wallet	clock (CP+Sem)	broom
39	for Valentine's day Jon is going to buy his girlfriend a bouquet of **flowers** and some chocolates.	flag	vase (CP+Sem)	pen
40	Before sitting down Jane had to pull up a **chair** and find space at the table.	cherry	desk (CP+Sem)	sword
41	The screen froze again so Laura decided to reboot the **computer** after waiting for half an hour.	compass	keyboard (CP+Sem)	flip flop

Table continues on the next page

Table C.1: Experimental stimuli — continued

#	Sentence	Phon. Comp.	Semantic Competitor	Unrelated
42	The boat gracefully drifted down the **river** as the people watched.	ribbon	fishing rod (CP+Sem)	star
43	Andria couldn't see well because the sun was in her **eyes** while she was driving.	ipad	mascara (CP-only)	cake platter
44	That dog looks so happy wagging its **tail** as it walks along.	table	paw (CP+Sem)	daisy
45	It had been raining heavily and Georges' boots were covered in **mud** from his long walk.	money	shovel (CP+Sem)	tie
46	Luke went to the library on Thursday but only borrowed one **book** for his assignment.	bulldozer	shelf (CP+Sem)	medal
47	Mark tried to wrap up his present but couldn't cut the paper with the blunt **scissors** very easily.	cymbal	cleaver (CP+Sem)	perfume
48	The man was gathering honey when he was stung by a **bee** and gave a cry.	beer	honeycomb (CP+Sem)	weight
49	Because Peter wants to travel off peak taking a coach would be more expensive than taking a **train** to London.	tray	dump truck (CP+Sem)	bench

Table continues on the next page

Table C.1: Experimental stimuli — continued

#	Sentence	Phon. Comp.	Semantic Competitor	Unrelated
50	Rory went to the opticians last week and chose a new pair of **glasses** for himself.	gloves	binoculars (CP+Sem)	flute
51	Helen was hungry so she made herself a bacon sandwich with a fried **egg** and chips.	elbow	hen (CP+Sem)	parachute
52	The runners are not supposed to start until the judge blows a **whistle** at the beginning of the race.	whisk	bell (CP-only)	helicopter
53	The screen turned itself off because the laptop had a flat **battery** again.	barrel	adapter (CP+Sem)	tomato
54	The woman couldn't find a peeler in the kitchen so she pealed the apple using a **knife** from the drawer.	knee	saw (CP-only)	pillow
55	The exam involves lots of complex multiplication and division so students are allowed to use a **calculator** for any of the questions.	calendar	scale (CP-only)	bullet
56	To make sushi the chef went to the market to buy some **fish** early in the morning.	finger	can (CP+Sem)	road

End of table

C.2 Experimental Sentence Transcriptions - Experiment 2

Table C.2: Phonetic transcriptions of experimental stimuli.

#	Phonetic Transcription
1	[ðəkomˈpjutəˈpɔɪntə ˈdʌzntmuːv ɑːlˈðəʊtɔmːi izˈmuːvɪŋ ðəmaʊzˈkwɪkli]
2	[ðəˈtʊərɪsts ɪkˈspektɪdreɪn wenðəsʌn wentbɪˈhaɪndðəklaʊds bʌtðəˈweðə gɑtˈbetəˈleɪtə]
3	[æˈmiːlɪə gɑteɪˈdraɪvɪŋˈlaɪsns əndwɪlbaɪhəəʊnkɑːtəˈmɑrəʊ]
4	[ˈdɪlən gɑtlɑstˈtədeɪ səʊtəˈmɑrəʊ hiːwɪljuːzhizmæp təfaɪndeɪweɪ]
5	*Item removed during stimulus cleaning.*
6	[ætðətreɪnˈsteɪʃən ðəkəˈmjuːtəzɑːˈrʌʃɪŋ təbaɪeɪˈtɪkɪt bɪˈfɔːðətreɪn liːvz]
7	[ðəˈrestrɑnt ɪzˈɔːlweɪzˈbɪzɪ səʊliːəʊ wilbʊkeɪˈteɪbəl fɔːðəhəʊl ˈfæmlɪ]
8	[keɪt ˈdɪdntlaɪkˈkofi səʊʃiːˈɔːdəd əkʌpoftiː wɪθhɜːkeɪk]
9	[megwɪlgəʊtəðəpɑːk təwɔːkhədɑg ðisɑːftəˈnuːn]
10	[ðəməʊl wɑzˈdɪgɪŋeɪsmɔːlhəʊl ɪnðəbækjɑːd]
11	[ˈhɪərɪŋðənɔɪzaʊtsaɪdðəˈklɑːsrʊm ðəˈlektʃərəkləʊzdðədɔː səʊðætðəˈstjuːdənts kænˈkɑnsəntreɪtbetə]
12	[ˈpiːpəlsɔː ðəfɜːstˈspeɪsʃɪp ðætˈlændɪd ɑnðəmuːn ɪnðəspeɪsmjuːˈziːəm]
13	[ˈmaɪəgɑtkəʊld əndhæzeɪˈrʌnɪnəʊz əndsəʊdəzhɜːˈbrʌðə]
14	[rəʊzkʊdəntiːtˈnuːdəlz ˈjuːzɪŋˈtʃɑpstɪks səʊjuːzdeɪfɔːkɪnˈsted]
15	[wenɪtsˈsʌnɪ seərəˈɔːlweɪzˈkʌvəzhɜːhed əndˈnɔːməlɪweəz eɪˈbeɪsbɔːlkæp aʊtˈsaɪd]
16	*Item removed during stimulus cleaning.*
17	[fɔːˈkrɪsməsðəˈtʃɪldrənɑːˈhæŋgɪŋbelzɑnðətriː əndˈpʊtɪŋðəˈstɑkɪŋz baɪðəfaɪə]
18	[təˈfɪnɪʃðəkeɪk ˈskɑːlɪtspred ðəwɪtkriːmˈsmuːðlɪ ˈəʊvəðətɑp]
19	[səʊʃiːkʊdrʌbaʊt enɪmɪˈsteɪksɪnhənəʊts beθdɪsaɪdɪdtəraɪtwɪθeɪˈpensəl fɔːɔːlhəklɑːsɪzðɪstɜːm]
20	[ˈhærɪ ɪnˈtendz təprəˈpəʊztəˈemɪlɪ əndgɪvhəðərɪŋ ɑnˈsætədɪ]
21	[æzeɪlaɪftaɪmvedʒɪˈteərɪən ɑˈlɪvɪə dʌsntmɪziːtɪŋmiːt ætɔːl]
22	[ðəɪksˈpensɪvwaɪn ɪzmeɪdfrɑmˈspeʃəltaɪpɑvgreɪp ðætɪzgrəʊnəʊnlɪ ɪnðəsaʊθɑvfrɑːns]
23	[ɑndðəˈaɪlənd dʒɪmwentˈswɪmɪŋ əndgɑteɪˈsʌntænætðəbiːtʃ ˈɑːftədʒʌsttuːaʊəz]
24	[ðətʃaɪldbɪˈliːvdðætˈsæntəklɔːz wədˈkʌmɪntuːhəhaʊz daʊnðəˈtʃɪmnɪætˈmɪdnaɪt]
25	[ðəˈtrævələwenttəðəˈdezət bɪˈkɑz hiːˈwɑntɪdtəraɪdeɪˈkæməl əndgəʊɪkˈsplɔːrɪŋ]
26	[kɪmfaʊndðəruːm wɑztuːhɑtəndˈhjuːmɪd səʊ təgetsʌmfreʃeə ʃɪːˈəʊpəndðəˈwɪndəʊˈfʊlɪ]

Table continues on the next page

Table C.2: Experimental sentence transcriptions — continued

#	**Phonetic Transcription**
27	[ɪnˈstedɑvˈsendɪŋænˈiːmeɪl ˈelɪ dɪˈsaɪdidtəraɪt eɪˈletə təhəˈsɪstə]
28	[ɪnænɪˈmɜːdʒənsɪ wiːˈkænɑtjuːzeɪlɪft ɪnˈsted wiːniːdtəjuːzðəsteəzfɔːˈaʊəˈseɪftɪ]
29	[ˈhænəbrɔːtəˈkæləndə əndhʌŋɪtɑnðəwɔːl neksttəhədesk]
30	[ðəkæmpˈliːdə tɔːttˈʃɪldrən haʊtəpɪtʃ eɪtent ɑnðeəəʊn]
31	[ɔːlðəɪmˈplɔɪəzhɪə hævˈseɪvɪŋəkaʊnts ætðəseɪmbæŋk fɔːsʌmˈriːzən]
32	[luːɪslɑsthɪzˈmemrɪ bɪˈkɑzɑvðəˈdæmɪdʒtəhɪzbreɪn frɑmhɪzstrəʊk]
33	[ˈvaɪələtleftðəˈdɜːtɪpleɪts əndkʌps ɪnðəsɪŋk ɔːldeɪ]
34	[ɪfjuːpʊtˈwɔːtəɪnðəˈfriːzə ɪtwɪlbɪˈkʌmaɪs wɪˈðɪneɪfjuːaʊəz]
35	[ɪnðəˈtenɪsˈlesən lɔːrənhɪtðəbɔːl wɪθhəˈrækɪt ˈverɪhɑːd]
36	[ˈkærənkʊdnɑtgetðəˈvɑlɪbɔːlʃɑt ˈəʊvəðənet æzʃiːwɑzsmɔːl]
37	[ðəwebhædbiːnspʌn baieɪlɑːdʒˈspaɪdə əndɪthædɔːlˈredɪkɔːt eɪlɑtɑvˈɪnsekts]
38	[sæmdɪdntnəʊðətaɪm bɪˈkɑzhɪːfəˈgɑttəweə hɪzwɑtʃtəˈdeɪ]
39	[fɔːˈvæləntaɪnzdeɪ dʒɑnɪzˈgəʊɪŋ təbaɪhɪzˈgɜːlfrend eɪbəʊˈkeɪɑvflaʊəz əndsʌmˈtʃɑkləts]
40	[bɪˈfɔːˈsɪtɪŋdaʊn dʒeɪnhædtəpʊlʌpeɪtʃeə əndfaɪndspeɪsætðəˈteɪbəl]
41	[ðəskriːnfrəʊzəˈgen səʊˈlɔːrədɪˈsaɪdɪd təriːˈbuːtðəkəmˈpjuːtə ˈɑːftəˈweɪtɪŋfɔːhɑːfænaʊə]
42	[ðəbəʊt ˈgreɪsfəlɪˈdrɪftɪd daʊnðəˈrɪvə æzðəˈpiːpəlwɑtʃt]
43	[ˈændrɪəˈkʊdəntsiːwel bɪˈkɑzðəsʌn wɑzɪnhəaɪz waɪlʃiːwɑzˈdraɪvɪŋ]
44	[ðætdɑglʊkssəʊˈhæpɪ ˈwæagɪŋɪtsteɪl æzɪtwɔːksəˈlɑn]
45	[ɪthædbiːnˈreɪnɪŋˈhevɪlɪ əndˈdʒɔːdʒɪzbuːts wɜːˈkʌvədɪnmʌd frɑmhɪzlɑŋwɔːk]
46	[luːkwenttəðəˈlaɪbrərɪɑnˈθɜːzdeɪ bʌtəʊnlɪˈbɑrəʊd wʌnbʊk fɔːhɪzəˈsaɪnmənt]
47	[mɑːktraɪdtəræpʌphɪzˈprezənt bʌtˈkʊdəntkʌtðəˈpeɪpə wɪθðəblʌntˈsɪzəzˈverɪˈiːzɪlɪ]
48	[ðəmænwɑzˈgæðərɪŋhʌnɪ wenhiːwɑzstʌŋ baɪeɪbiː əndgeɪveɪkraɪ]
49	[bɪˈkɑzˈpiːtəwɑntstəˈtrævəlɑfpiːk ˈteɪkɪŋeɪkəʊtʃ wədbiːmɔːɪksˈpensɪv ðənˈteɪkɪŋeɪtreɪn təˈlʌndən]
50	[ˈrərɪwenttəðəɑpˈtɪʃənzlɑːstwiːk əndtʃəʊz eɪnjuːpeəɑvˈglɑːsɪz fɔːhɪmˈself]
51	[ˈhelənwɑzˈhʌŋgrɪ səʊʃiːmeɪdhɜːˈself eɪˈbeɪkənˈsænwɪtʃ wɪθeɪfraɪdegəndtʃɪps]
52	[ðəˈrʌnəz ɑːnɑtsəˈpəʊzd təstɑːt ʌntɪlðədʒʌdʒbləʊzaɪwɪsəl ætðəbɪˈgɪnɪŋɑvðəreɪs]
53	[ðəskriːn tɜːndɪtˈselfɑf biˈkɑzðəˈlæptɑp hædeɪflætˈbætrɪəˈgen]
54	[ðəˈwʊmənˈkʊdəntfaɪndeɪˈpiːləɪnðəˈkɪtʃən səʊʃiːpiːldðiˈæpəl ˈjuːzɪŋeɪnaɪffrɑmðədrɔː]
55	[ðəɪgˈzæmɪnˈvɑlvz lɑtsɑvˈkɑmpleksmʌltɪplɪˈkeɪʃənənddɪˈvɪʒən səʊˈstjuːdəntsɑːəˈlaʊd təjuːz aˈkælkjʊleɪtə fɔːenɪɑvðəˈkwestʃənz]

Table continues on the next page

Table C.2: Experimental sentence transcriptions — continued

#	**Phonetic Transcription**
56	[təmeɪkˈsuːʃɪ ðəʃefwenttəðəˈmɑːkɪt təbaɪsʌmfɪʃ ˈɜːlɪɪnðəˈmɔːnɪŋ]

End of table

Appendix D

Statistical Model Specifications. Experiment 1a

D.1 GAMM Model Output for Anticipatory Window

Table D.1: GAMM results (anticipatory window), Part A: Parametric coefficients (Experiment 1a).

Parametric coefficients (fixed effects)

Term	Estimate	Std. Error	t value	p
Intercept	-0.80	0.07	-11.42	$< .0001$
phonological.L1	0.11	0.09	1.20	.23
target.L1	1.07	0.11	9.36	$< .0001$
control.L2	-0.08	0.10	-0.80	.42
phonological.L2	0.00	0.10	0.01	.99
target.L2	0.56	0.11	5.02	$< .0001$
control.Dyslexia	-0.11	0.12	-0.90	.37
phonological.Dyslexia	0.13	0.12	1.07	.28
target.Dyslexia	0.89	0.17	5.38	$< .0001$

Table D.2: GAMM results (anticipatory window), Part B: Approximate significance of smooth terms (Experiment 1a).

Approximate significance of smooth terms

Smooth term	edf	Ref. df	*F*	*p*
s(time)	4.37	5.54	1.51	.19
s(time):phonological.L1	1.00	1.00	0.11	.74
s(time):target.L1	1.00	1.00	13.17	< .001
s(time):control.L2	1.00	1.00	0.04	.84
s(time):phonological.L2	4.25	4.85	0.78	.59
s(time):target.L2	1.00	1.00	3.47	.06
s(time):control.Dyslexia	1.00	1.00	0.01	.94
s(time):phonological.Dyslexia	1.00	1.00	0.27	.61
s(time):target.Dyslexia	1.00	1.00	4.55	.03
s(time, participant):phonological.L1	133.81	206.00	2.85	< .0001
s(time, participant):target.L1	127.11	206.00	2.04	< .0001
s(time, participant):control.L2	95.45	170.00	1.88	< .0001
s(time, participant):phonological.L2	98.37	170.00	2.07	< .0001
s(time, participant):target.L2	106.66	170.00	2.21	< .0001
s(time, participant):control.Dyslexia	57.05	107.00	2.07	< .0001

Table continues on the next page

Table D.2: GAMM results (anticipatory window), Part B continued: Smooth terms

Approximate significance of smooth terms — continued

Smooth term	edf	Ref. df	*F*	*p*
s(time, participant):phonological.Dyslexia	59.05	107.00	2.11	$< .0001$
s(time, participant):target.Dyslexia	69.72	107.00	2.45	$< .0001$
s(time, item)	133.49	188.00	3.13	$< .0001$

Notes. For parametric coefficients, *Estimate* is the effect size, *Std. Error* its standard error, t the test statistic, and p the significance level. For smooth terms, *edf* is the estimated degrees of freedom, *Ref. df* the reference degrees of freedom, F the test statistic, and p the significance level. Smooths `s(time)` are population-level; `s(time, participant)` and `s(time, item)` are random-effect smooths.

D.2 GAMM Model Output for Resolution Window

Table D.3: GAMM results (resolution window), Part A: Parametric coefficients (Experiment 1a).

Parametric coefficients (fixed effects)				
Term	Estimate	Std. Error	t value	p
Intercept	-0.78	0.07	-11.13	$< .0001$
phonological.L1	0.08	0.09	0.95	.34
target.L1	0.70	0.11	6.43	$< .0001$
control.L2	-0.05	0.10	-0.50	.62
phonological.L2	-0.01	0.10	-0.11	.91
target.L2	0.33	0.11	2.90	.004
control.Dyslexia	-0.08	0.12	-0.70	.49
phonological.Dyslexia	0.09	0.12	0.77	.44
target.Dyslexia	0.45	0.17	2.67	.008

Table D.4: GAMM results (resolution window), Part B: Approximate significance of smooth terms (Experiment 1a).

Approximate significance of smooth terms

Smooth term	edf	Ref. df	*F*	*p*
s(time)	3.83	5.04	1.48	.20
s(time):phonological.L1	1.00	1.00	0.00	.95
s(time):target.L1	1.00	1.00	9.89	.002
s(time):control.L2	1.00	1.00	0.10	.75
s(time):phonological.L2	3.75	4.35	0.76	.57
s(time):target.L2	1.00	1.00	1.38	.24
s(time):control.Dyslexia	1.00	1.00	0.01	.93
s(time):phonological.Dyslexia	1.00	1.00	0.04	.84
s(time):target.Dyslexia	1.00	1.00	1.46	.23
s(time, participant):phonological.L1	123.09	206.00	2.36	$< .0001$
s(time, participant):target.L1	115.45	206.00	1.74	$< .0001$
s(time, participant):control.L2	90.31	170.00	1.70	$< .0001$
s(time, participant):phonological.L2	95.49	170.00	1.89	$< .0001$
s(time, participant):target.L2	102.53	170.00	1.98	$< .0001$
s(time, participant):control.Dyslexia	50.89	107.00	1.66	$< .0001$

Table continues on the next page

Table D.4: GAMM results (resolution window), Part B continued: Smooth terms

Approximate significance of smooth terms — continued				
Smooth term	edf	Ref. df	F	p
s(time, participant):phonological.Dyslexia	50.41	107.00	1.34	$< .0001$
s(time, participant):target.Dyslexia	10.37	107.00	0.30	
s(time, item)	79.04	188.00	1.13	$< .0001$

Notes. As above for definitions of columns and smooths.

Appendix E

Statistical Model Specifications. Experiment 1b

E.1 GAMM Model Output for Resolution Window

Table E.1: GAMM results (resolution window), Part A: Parametric coefficients (Experiment 1b).

Parametric coefficients (fixed effects)				
Term	Estimate	Std. Error	t value	p
Intercept	0.40	0.16	2.53	.01
congruent.L1	0.60	0.24	2.56	.01
incongruent.L1	0.32	0.25	1.27	.21
control.L2	-0.35	0.12	-2.80	.01
congruent.L2	0.35	0.24	1.48	.14
incongruent.L2	-0.07	0.26	-0.27	.79
control.Dyslexia	0.03	0.23	0.15	.88
congruent.Dyslexia	0.35	0.29	1.21	.23
incongruent.Dyslexia	0.41	0.26	1.62	.10

Table E.2: GAMM results (resolution window), Part B: Approximate significance of smooth terms (Experiment 1b).

Approximate significance of smooth terms

Smooth term	edf	Ref. df	*F*	*p*
s(time)	5.13	6.10	27.49	< .0001
s(time):congruent.L1	2.88	3.43	3.06	.02
s(time):incongruent.L1	1.00	1.00	0.20	.65
s(time):control.L2	2.41	2.82	1.88	.11
s(time):congruent.L2	1.20	1.30	0.91	.45
s(time):incongruent.L2	2.14	2.50	1.72	.17
s(time):control.Dyslexia	1.00	1.00	6.71	.01
s(time):congruent.Dyslexia	1.00	1.00	0.26	.61
s(time):incongruent.Dyslexia	1.00	1.00	1.06	.30
`s(time, participant):congruent.L1`	59.23	206.00	1.08	< .0001
`s(time, participant):incongruent.L1`	66.20	206.00	2.16	< .0001
`s(time, participant):control.L2`	65.16	176.00	1.55	< .0001
`s(time, participant):congruent.L2`	36.59	177.00	0.78	< .0001
`s(time, participant):incongruent.L2`	59.20	179.00	2.09	< .0001
`s(time, participant):control.Dyslexia`	31.80	105.00	2.14	< .0001

Table continues on the next page

Table E.2: GAMM results (resolution window), Part B continued: Smooth terms

Approximate significance of smooth terms — continued				
Smooth term	edf	Ref. df	F	p
`s(time, participant):congruent.Dyslexia`	27.85	106.00	1.20	$< .0001$
`s(time, participant):incongruent.Dyslexia`	39.24	106.00	1.33	$< .0001$
`s(time, item)`	68.06	159.00	5.62	$< .0001$

Notes. As above for definitions of columns and smooths.

E.2 GAMM Model Output for Resolution Window (Combined Conditions)

Table E.3: GAMM results (resolution window, combined conditions), Part A: Parametric coefficients (Experiment 1b).

Parametric coefficients (fixed effects)				
Term	Estimate	Std. Error	t value	p
Intercept	0.37	0.17	2.18	.03
predictive.Italian	0.48	0.22	2.20	.03
control.Dyslexia	-0.03	0.23	-0.14	.89
predictive.Dyslexia	0.39	0.24	1.58	.11
control.Tyrolean	-0.37	0.13	-2.90	.004
predictive.Tyrolean	0.14	0.23	0.63	.53

Table E.4: GAMM results (resolution window, combined conditions), Part B: Approximate significance of smooth terms (Experiment 1b).

Approximate significance of smooth terms

Smooth term	edf	Ref. df	*F*	*p*
s(time)	6.67	7.74	17.08	< .0001
s(time):predictive.Italian	3.18	3.72	2.24	.07
s(time):control.Dyslexia	1.00	1.00	5.73	.02
s(time):predictive.Dyslexia	1.68	1.90	1.03	.43
s(time):control.Tyrolean	1.63	1.83	0.72	.37
s(time):predictive.Tyrolean	1.00	1.00	1.18	.28
`s(time, participant):predictive.Italian`	100.79	206.00	1.26	< .0001
`s(time, participant):control.Dyslexia`	51.31	105.00	1.45	< .0001
`s(time, participant):predictive.Dyslexia`	54.19	107.00	1.32	< .0001
`s(time, participant):control.Tyrolean`	97.74	176.00	1.72	< .0001
`s(time, participant):predictive.Tyrolean`	79.49	179.00	1.03	< .0001
`s(time, item)`	109.91	160.00	3.44	< .0001

Notes. As above for definitions of columns and smooths.

Appendix F

Statistical Model Specifications. Experiment 2

F.1 GAMM Model Output for Anticipatory Window (Experiment 2)

Table F.1: GAMM results (anticipatory window), Part A: Parametric coefficients (Experiment 2).

Parametric coefficients (fixed effects)

Term	Estimate	Std. Error	t value	p
Intercept	-0.72	0.03	-26.99	$< .0001$
Target.L1	0.67	0.06	12.06	$< .0001$
Phonological.L1	0.09	0.04	2.47	.014
Semantic.L1	0.28	0.04	6.34	$< .0001$
Unrelated.L1dys	0.08	0.04	2.16	.031
Target.L1dys	0.80	0.06	12.60	$< .0001$
Phonological.L1dys	0.10	0.04	2.75	.006
Semantic.L1dys	0.32	0.04	7.91	$< .0001$
Unrelated.L2	0.04	0.04	1.09	.28
Target.L2	0.71	0.07	9.48	$< .0001$
Phonological.L2	0.09	0.04	2.60	.009
Semantic.L2	0.30	0.04	6.96	$< .0001$

Table F.2: GAMM results (anticipatory window), Part B: Approximate significance of smooth terms (Experiment 2).

Approximate significance of smooth terms

Smooth term	edf	Ref. df	F	p
s(time)	1.00	1.00	0.21	.65
s(time):Target.L1	9.05	10.73	2.64	.004
s(time):Phonological.L1	1.00	1.00	0.02	.90
s(time):Semantic.L1	1.00	1.00	2.55	.11
s(time):Unrelated.L1dys	1.00	1.00	0.10	.76
s(time):Target.L1dys	7.37	8.97	4.26	$< .0001$
s(time):Phonological.L1dys	1.00	1.00	0.02	.89
s(time):Semantic.L1dys	7.76	9.43	1.06	.50
s(time):Unrelated.L2	1.00	1.00	0.02	.90
s(time):Target.L2	1.00	1.00	6.39	.011
s(time):Phonological.L2	1.00	1.00	0.19	.66
s(time):Semantic.L2	10.56	12.12	2.59	.002
`s(time, participant):Target.L1`	266.06	350.00	3.31	$< .0001$
`s(time, participant):Phonological.L1`	219.11	350.00	1.69	$< .0001$
`s(time, participant):Semantic.L1`	258.17	350.00	2.79	$< .0001$

Table continues on the next page

Table F.2: GAMM results (anticipatory window), Part B continued: Smooth terms

Approximate significance of smooth terms — continued

Smooth term	edf	Ref. df	F	p
`s(time, participant):Unrelated.L1dys`	176.34	260.00	2.13	$< .0001$
`s(time, participant):Target.L1dys`	192.71	260.00	3.01	$< .0001$
`s(time, participant):Phonological.L1dys`	175.30	260.00	2.18	$< .0001$
`s(time, participant):Semantic.L1dys`	175.09	260.00	2.18	$< .0001$
`s(time, participant):Unrelated.L2`	138.45	215.00	2.12	$< .0001$
`s(time, participant):Target.L2`	162.36	215.00	3.27	$< .0001$
`s(time, participant):Phonological.L2`	132.64	215.00	2.00	$< .0001$
`s(time, participant):Semantic.L2`	146.01	215.00	2.49	$< .0001$
`s(time, item)`	366.00	485.00	3.57	$< .0001$

Notes. As above for definitions of columns and smooths.

F.2 GAMM Model Output for Anticipatory Window (Cp+Sem vs. Cp-only; Experiment 2)

Table F.3: GAMM results (anticipatory window), Part A: Parametric coefficients (Cp+Sem vs. Cp-only; Experiment 2).

Parametric coefficients (fixed effects)				
Term	Estimate	Std. Error	t value	p
Intercept	-0.74	0.03	-25.27	$< .0001$
CpSem.L1	0.50	0.06	7.81	$< .0001$
CpOnly.L1	0.05	0.05	0.90	.37
Unrelated.L1dys	0.08	0.04	2.13	.033
CpSem.L1dys	0.45	0.06	7.44	$< .0001$
CpOnly.L1dys	0.11	0.05	2.11	.035
Unrelated.L2	0.06	0.04	1.58	.11
CpSem.L2	0.44	0.09	5.11	$< .0001$
CpOnly.L2	0.21	0.07	2.98	.003

Table F.4: GAMM results (anticipatory window), Part B: Approximate significance of smooth terms (Experiment 2).

Approximate significance of smooth terms

Smooth term	edf	Ref. df	*F*	*p*
s(time)	1.94	2.18	0.46	.67
s(time):CpSem.L1	1.71	1.92	1.14	.39
s(time):CpOnly.L1	1.00	1.00	0.03	.85
s(time):Unrelated.L1dys	1.00	1.00	0.39	.53
s(time):CpSem.L1dys	9.88	11.52	2.51	.004
s(time):CpOnly.L1dys	1.30	1.38	0.37	.73
s(time):Unrelated.L2	2.49	2.92	0.44	.67
s(time):CpSem.L2	9.76	11.46	2.28	.007
s(time):CpOnly.L2	2.28	2.62	0.71	.60
`s(time, participant):CpSem.L1`	271.77	350.00	3.63	$< .0001$
`s(time, participant):CpOnly.L1`	207.72	314.00	2.68	$< .0001$
`s(time, participant):Unrelated.L1dys`	174.40	260.00	2.55	$< .0001$
`s(time, participant):CpSem.L1dys`	186.61	260.00	3.04	$< .0001$
`s(time, participant):CpOnly.L1dys`	175.51	260.00	2.83	$< .0001$
`s(time, participant):Unrelated.L2`	125.43	215.00	1.69	$< .0001$

Table continues on the next page

Table F.4: GAMM results (anticipatory window), Part B continued: Smooth terms

Approximate significance of smooth terms — continued

Smooth term	edf	Ref. df	F	p
`s(time, participant):CpSem.L2`	147.99	215.00	2.72	$< .0001$
`s(time, participant):CpOnly.L2`	151.77	215.00	3.19	$< .0001$
`s(time, item)`	360.37	485.00	3.20	$< .0001$

Notes. As above for definitions of columns and smooths.

F.3 GAMM Model Output for Resolution Window (Experiment 2, Predictable Items)

Table F.5: GAMM results (resolution window), Part A: Parametric coefficients (Experiment 2, predictable items).

Parametric coefficients (fixed effects)				
Term	Estimate	Std. Error	t value	p
Intercept	-0.73	0.04	-20.34	$< .0001$
Target.L1	1.43	0.07	20.93	$< .0001$
Phonological.L1	-0.02	0.05	-0.51	.61
Semantic.L1	0.45	0.07	5.96	$< .0001$
Unrelated.L1dys	0.00	0.07	0.04	.97
Target.L1dys	1.55	0.07	21.90	$< .0001$
Phonological.L1dys	0.14	0.06	2.13	.033
Semantic.L1dys	0.48	0.07	7.20	$< .0001$
Unrelated.L2	-0.01	0.06	-0.21	.83
Target.L2	1.35	0.10	13.98	$< .0001$
Phonological.L2	0.05	0.06	0.82	.41
Semantic.L2	0.58	0.08	7.26	$< .0001$

Table F.6: GAMM results (resolution window), Part B: Approximate significance of smooth terms (Experiment 2, predictable items).

Approximate significance of smooth terms

Smooth term	edf	Ref. df	F	p
s(time)	3.26	3.82	2.06	.063
s(time):Target.L1	1.00	1.00	4.15	.042
s(time):Phonological.L1	1.00	1.00	0.01	.93
s(time):Semantic.L1	1.01	1.01	0.01	.94
s(time):Unrelated.L1dys	2.26	2.61	0.85	.37
s(time):Target.L1dys	3.43	4.14	2.23	.074
s(time):Phonological.L1dys	2.30	2.68	1.63	.24
s(time):Semantic.L1dys	1.00	1.00	2.05	.15
s(time):Unrelated.L2	1.00	1.00	0.08	.77
s(time):Target.L2	2.47	3.05	4.49	.003
s(time):Phonological.L2	1.00	1.00	0.26	.61
s(time):Semantic.L2	3.86	4.55	3.14	.010
`s(time, participant):Target.L1`	111.70	350.00	0.78	$< .0001$
`s(time, participant):Phonological.L1`	111.30	350.00	0.60	$< .0001$
`s(time, participant):Semantic.L1`	179.43	350.00	1.76	$< .0001$

Table continues on the next page

Table F.6: GAMM results (resolution window), Part B continued: Smooth terms

Approximate significance of smooth terms — continued

Smooth term	edf	Ref. df	F	p
`s(time, participant):Unrelated.L1dys`	128.57	260.00	1.40	$< .0001$
`s(time, participant):Target.L1dys`	101.60	260.00	0.98	$< .0001$
`s(time, participant):Phonological.L1dys`	123.28	260.00	1.22	$< .0001$
`s(time, participant):Semantic.L1dys`	135.92	260.00	1.48	$< .0001$
`s(time, participant):Unrelated.L2`	72.22	215.00	0.66	$< .0001$
`s(time, participant):Target.L2`	65.01	215.00	0.92	$< .0001$
`s(time, participant):Phonological.L2`	88.01	215.00	1.02	$< .0001$
`s(time, participant):Semantic.L2`	100.71	215.00	1.30	$< .0001$
`s(time, item)`	202.92	485.00	1.25	$< .0001$

Notes. As above for definitions of columns and smooths.

F.4 GAMM Model Output for Resolution Window, predictable items (Cp+Sem vs. Cp-only; Experiment 2)

Table F.7: GAMM results (resolution window), Part A: Parametric coefficients (Cp+Sem vs. Cp-only; Experiment 2).

Parametric coefficients (fixed effects)				
Term	Estimate	Std. Error	t value	p
Intercept	-0.74	0.04	-17.55	$< .0001$
CpSem.L1	0.55	0.08	6.67	$< .0001$
CpOnly.L1	0.19	0.10	1.88	.06
Unrelated.L1dys	0.01	0.07	0.18	.86
CpSem.L1dys	0.57	0.08	6.96	$< .0001$
CpOnly.L1dys	0.29	0.09	3.04	.002
Unrelated.L2	-0.01	0.06	-0.16	.87
CpSem.L2	0.68	0.09	7.43	$< .0001$
CpOnly.L2	0.37	0.15	2.53	.01

Table F.8: GAMM results (resolution window), Part B: Approximate significance of smooth terms (Experiment 2).

Approximate significance of smooth terms

Smooth term	edf	Ref. df	*F*	*p*
s(time)	3.60	4.23	2.16	.07
s(time):CpSem.L1	1.00	1.01	0.02	.90
s(time):CpOnly.L1	1.00	1.00	0.23	.63
s(time):Unrelated.L1dys	2.34	2.75	0.75	.40
s(time):CpSem.L1dys	1.00	1.00	2.65	.10
s(time):CpOnly.L1dys	2.62	3.04	0.90	.41
s(time):Unrelated.L2	1.00	1.00	0.11	.74
s(time):CpSem.L2	3.38	4.00	2.61	.03
s(time):CpOnly.L2	1.00	1.00	2.75	.10
`s(time, participant):CpSem.L1`	165.42	350.00	1.49	$< .0001$
`s(time, participant):CpOnly.L1`	116.59	313.00	0.98	$< .0001$
`s(time, participant):Unrelated.L1dys`	117.67	260.00	1.20	$< .0001$
`s(time, participant):CpSem.L1dys`	114.82	260.00	1.12	$< .0001$
`s(time, participant):CpOnly.L1dys`	117.60	260.00	1.27	$< .0001$
`s(time, participant):Unrelated.L2`	64.98	215.00	0.56	$< .0001$

Table continues on the next page

Table F.8: GAMM results (resolution window), Part B continued: Smooth terms

Approximate significance of smooth terms — continued				
Smooth term	edf	Ref. df	F	p
`s(time, participant):CpSem.L2`	97.60	215.00	1.26	$< .0001$
`s(time, participant):CpOnly.L2`	58.07	214.00	0.77	$< .0001$
`s(time, item)`	210.12	485.00	1.30	$< .0001$

Notes. As above for definitions of columns and smooths.

F.5 GAMM Model Output for Resolution Window (Unpredictable Items, Experiment 2)

Table F.9: GAMM results (resolution window, unpredictable items), Part A: Parametric coefficients (Experiment 2).

Parametric coefficients (fixed effects)				
Term	Estimate	Std. Error	t value	p
Intercept	-0.75	0.03	-22.27	$< .0001$
Target.L1	1.05	0.04	24.70	$< .0001$
Phonological.L1	0.21	0.04	4.76	$< .0001$
Semantic.L1	0.32	0.05	7.01	$< .0001$
Unrelated.L1dys	0.02	0.04	0.56	.58
Target.L1dys	1.06	0.05	20.18	$< .0001$
Phonological.L1dys	0.25	0.05	4.61	$< .0001$
Semantic.L1dys	0.37	0.06	6.66	$< .0001$
Unrelated.L2	-0.02	0.04	-0.45	.65
Target.L2	1.00	0.05	18.46	$< .0001$
Phonological.L2	0.29	0.05	6.13	$< .0001$
Semantic.L2	0.23	0.05	4.36	$< .0001$

Table F.10: GAMM results (resolution window, unpredictable items), Part B: Approximate significance of smooth terms (Experiment 2).

Approximate significance of smooth terms

Smooth term	edf	Ref. df	F	p
s(time)	6.15	7.14	3.61	.002
s(time):Target.L1	6.21	7.38	27.73	$< .0001$
s(time):Phonological.L1	5.31	6.26	4.21	$< .0001$
s(time):Semantic.L1	4.18	4.90	6.83	$< .0001$
s(time):Unrelated.L1dys	1.00	1.00	1.38	.24
s(time):Target.L1dys	6.71	7.98	23.08	$< .0001$
s(time):Phonological.L1dys	3.97	4.63	1.73	.13
s(time):Semantic.L1dys	3.46	3.98	3.48	.02
s(time):Unrelated.L2	4.34	5.20	1.90	.09
s(time):Target.L2	5.56	6.66	18.43	$< .0001$
s(time):Phonological.L2	1.00	1.00	4.46	.03
s(time):Semantic.L2	3.71	4.36	3.20	.02
s(time, participant):Target.L1	170.96	350.00	1.14	$< .0001$
s(time, participant):Phonological.L1	200.33	350.00	1.65	$< .0001$
s(time, participant):Semantic.L1	207.89	350.00	2.01	$< .0001$

Table continues on the next page

Table F.10: GAMM results (resolution window, unpredictable items), Part B continued: Smooth terms

Approximate significance of smooth terms — continued

Smooth term	edf	Ref. df	F	p
s(time, participant):Unrelated.L1dys	132.78	260.00	1.47	$< .0001$
s(time, participant):Target.L1dys	119.25	260.00	1.01	$< .0001$
s(time, participant):Phonological.L1dys	165.37	260.00	2.21	$< .0001$
s(time, participant):Semantic.L1dys	172.53	260.00	2.46	$< .0001$
s(time, participant):Unrelated.L2	126.49	215.00	1.54	$< .0001$
s(time, participant):Target.L2	141.44	215.00	2.38	$< .0001$
s(time, participant):Phonological.L2	121.71	215.00	1.92	$< .0001$
s(time, participant):Semantic.L2	128.57	215.00	2.09	$< .0001$
s(time, item)	415.91	485.00	3.88	$< .0001$

Notes. As above for definitions of columns and smooths.

Appendix G

Correlations. Experiment 1a

Table G.1: Correlations between mean divergence point to target and language background (LEAP-Q) variables for the Tyrolean L1 / Italian L2 group.

Variable	r	p
Daily Italian exposure	0.281	0.275
Self-rated L1 (Tyrolean) proficiency	0.076	0.757
Self-rated L2 (Italian) proficiency	0.191	0.434
Years of Italian use (school/work)	0.034	0.893
Age of Italian acquisition	–0.222	0.361
Daily L1 exposure	–0.061	0.815

Note. Pearson correlations between target divergence point and individual-difference variables assessed via the LEAP-Q. No correlation reached statistical significance ($p < .05$).

Appendix H

Correlations. Experiment 2

Table H.1: Correlations between CP+Sem divergence point and LEAP-Q variables for the L2 group.

Variable	r	p
Daily English exposure	0.143	0.530
Years in English-speaking country	0.527	0.008
Self-rated L1 proficiency	−0.271	0.200
Self-rated English proficiency	−0.260	0.220
Years of English use (school/work)	0.294	0.163
Age of English acquisition	0.217	0.309
Daily L1 exposure	−0.135	0.530

Note. Pearson correlations between CP+Sem divergence point and LEAP-Q individual-difference variables. No correlations reached statistical significance ($p < .05$).

Table H.2: Correlations between target divergence point and individual-difference measures across the full sample.

Variable	r	p
RAN Letters (rate)	−0.059	0.574
RAN Digits (rate)	−0.001	0.990
ARHQ total score	−0.063	0.550
ADHD overall score	−0.017	0.870
PPVT standard score	−0.006	0.953
ART score	−0.060	0.565
Stroop interference score	−0.014	0.893

Note. Pearson correlations between the divergence point for the target object and individual-difference measures. No correlations reached statistical significance ($p < .05$).

Table H.3: Correlations between semantic competitor divergence point and individual-difference measures across the full sample.

Variable	r	p
RAN Letters (rate)	0.054	0.605
RAN Digits (rate)	−0.008	0.937
ARHQ total score	0.009	0.931
ADHD overall score	0.005	0.960
PPVT standard score	−0.062	0.557
ART score	−0.018	0.861
Stroop interference score	−0.056	0.597

Note. Pearson correlations between the divergence point for the semantic competitor and individual-difference measures. No correlations reached statistical significance ($p < .05$).

Table H.4: Correlations between CP+Sem divergence point and individual-difference measures across the full sample.

Variable	r	p
RAN Letters (rate)	0.020	0.852
RAN Digits (rate)	−0.026	0.802
ARHQ total score	−0.007	0.947
ADHD overall score	0.046	0.664
PPVT standard score	−0.184	0.077
ART score	−0.132	0.207
Stroop interference score	0.123	0.241

Note. Pearson correlations between the divergence point for the CP+Sem competitor and individual-difference measures. No correlations reached statistical significance ($p < .05$).

Appendix I

Norming study materials. Experiment 2

Word Association Survey

1. You will see an image
2. Below the image will be a list of words
3. Select **ALL** words that you think are related to the image in any way:
    - The relationship can be based on meaning, function, typical situations, or common connections.
4. Click "Next" to proceed to the next image

There are no right or wrong answers - we're interested in your natural associations.

Figure I.1: Instruction screen shown to participants in the norming study for Experiment 2, describing the word–image association task.