Université de Montréal

Lexico-semantic and morphosyntactic processing in French-speaking adolescents with and without developmental language disorder

by Émilie Courteau

École d'orthophonie et d'audiologie, Faculté de médecine

Thesis submitted in partial fulfillment of the requirements for a degree in

Doctorat en sciences de l'orthophonie et de l'audiologie

Was evaluated by the following committee members:

Karine Marcotte, PhD (UdeM, president)

Phaedra Royle, PhD (UdeM, supervisor)

Karsten Steinhauer, PhD (McGill University, co-supervisor)

Natacha Trudeau, PhD (UdeM, internal member)

Michael T. Ullman, PhD (Georgetown University, external member)

March 2022

Résumé de la thèse

Bien que la communauté scientifique soit toujours à la recherche d'une caractéristique déterminante du trouble développemental du langage (TDL), les difficultés d'accord sujet-verbe, et par extension morphosyntaxiques, ont été identifiées comme un marqueur du TDL chez les enfants anglophones, autant chez les enfants du préscolaire que les plus vieux. Cependant, des études sur les enfants francophones d'âge préscolaire suggèrent que les déficits morphosyntaxiques ne seraient pas un marqueur fiable du TDL. Puisque que certains aspects de la morphosyntaxe en français ne sont acquis que vers l'âge de huit ans chez les enfants au développement typique, tels que l'accord en nombre des verbes sous-réguliers et irréguliers, ci-après SOUSIRR, les déficits morphosyntaxiques pourraient être un marqueur du TDL en français uniquement vers la (pré-)adolescence. Cette thèse a pour objectifs de déterminer si les (pré-)adolescents francophones au développement typique ont acquis l'accord en nombre des verbes SOUSIRR, si les (pré)adolescents francophones avec un TDL ont des déficits d'accord en nombre des verbes SOUSIRR, et à établir si la morphosyntaxe est un domaine de faiblesse par rapport à la lexico-sémantique dans cette population. L'accord en nombre des verbes SOUSIRR et les compétences morphosyntaxiques ont été évalués à l'aide de tâches ciblant les niveaux comportemental et neurocognitif en utilisant des tâches linguistiques et des potentiels évoqués (PÉ). De plus, nous avons développé des prédictions basées sur deux théories touchant les compétences morphosyntaxiques chez les (pré-)adolescents atteints de TDL : l'hypothèse du déficit procédural (Ullman & Pierpont, 2005 ; Ullman et al., 2020), et l'hypothèse du ralentissement généralisé (Kail, 1994). Cette thèse est composée de trois manuscrits pour publication. Le premier évalue les compétences des participants dans plusieurs domaines linguistiques, à l'aide de tâches comportementales typiquement utilisées en orthophonie et dans la recherche sur l'acquisition du langage. Les données révèlent des déficits lexico-sémantiques et morphosyntaxiques chez les participants avec un TDL, mais suggèrent qu'une tâche d'accord en nombre des verbes SOUSIRR était la meilleure pour discriminer les participants avec et sans TDL. Le deuxième article présente une étude innovante de PÉs utilisant uniquement des phrases grammaticales, présentées simultanément avec des images sémantiquement ou grammaticalement congruentes et incongruentes, afin d'évaluer le traitement morphosyntaxique et lexico-sémantique des phrases au niveau neurocognitif. Les résultats provenant de vingt-huit adultes francophones montrent qu'ils ont présenté les composantes PÉs attendues et comparables aux études utilisant des phrases agrammaticales. Ces données ont servi de référence pour établir si nos participants avec et sans TDL avaient un traitement linguistique mature. Le troisième article a testé cette nouvelle expérimentation avec nos participants (pré-)adolescents. Les résultats suggèrent que, contrairement à la morphosyntaxe, la lexico-sémantique est une force relative chez les adolescents avec un TDL lors du traitement de l'information linguistique au niveau neurocognitif. Dans l'ensemble, cette thèse révèle que la morphosyntaxe est particulièrement altérée chez les adolescents francophones avec un TDL. Nous discutons les résultats en relation avec la pratique clinique orthophonique et soulignons l'importance d'examiner les processus neurocognitifs dans l'étude du TDL.

Mots-clés : Adolescence, évaluation, trouble développemental du langage, potentiels évoqués, français, morphosyntaxe, neurocognition, sémantique, orthophonie, accord sujet-verbe.

General abstract

Although the scientific community is still searching for a defining characteristic of developmental language disorder (DLD), problems with subject-verb agreement, and by extension morphosyntax, have been identified as a hallmark of English-speaking preschoolers and older children with DLD. However, in studies of French-speaking preschoolers with DLD, morphosyntax has not been found to be a specific linguistic weakness. Since there is evidence that some aspects of morphosyntax in French are acquired by children with typical language (TL) development only later in childhood, such as subregular and irregular subject-verb number agreement, henceforth SUBIRR, morphosyntax has been argued to be a French marker for DLD only in older childhood and adolescence. The present thesis aimed to determine if French speaking (pre-)teenagers with TL have acquired SUBIRR number agreement, resolve whether French-speaking (pre-)teenagers with DLD are impaired on SUBIRR number agreement, and establish whether morphosyntax is an area of weakness as compared to lexico-semantics in this population. SUBIRR number agreement and morphosyntactic skills were evaluated with tasks targeting the behavioural and neurocognitive levels using linguistics tasks and event-related potentials (ERP). Furthermore, we contrasted two theories' predictions on morphosyntactic skills in (pre-)teens with DLD : the procedural deficit hypothesis (Ullman & Pierpont, 2005; Ullman et al., 2020), and the generalized slowing hypothesis (Kail, 1994). This thesis is composed of three manuscripts for publication. The first evaluated our participants' skills in multiple linguistic domains with behavioural tasks typical of clinical and research settings. Data reveal impairments in the DLD group in both lexico-semantic and morphosyntactic domains but suggest that a SUBIRR number agreement task was best at discriminating DLD from controls. The second article presents a novel ERP experimental design using only grammatical sentences, presented

simultaneously with semantically and grammatically congruent or incongruent images, to assess morphosyntactic and lexico-semantic sentence processing at the neurocognitive level. Data from twenty-eight French-speaking adults show that they elicited the expected ERP components found in previous studies using ungrammatical sentences. These data served as a reference to establish whether our participants with and without TL process sentences in a mature way. The third article tested this novel ERP experiment with our (pre-)teen participants. We tested predictions of the procedural deficit hypothesis which states that children with DLD should have impaired morphosyntax due to an underlying procedural memory deficit, and the generalized slowing hypothesis, which proposes that all linguistic domains should be impaired due to an underlying processing deficit. This experimental design was run on teens with and without DLD. Although some processing delays were found in the DLD group, results on most conditions better fit the procedural deficit hypothesis. This study suggests that, in contrast with morphosyntax, lexico-semantics is a relative strength in teenagers with DLD when processing linguistic information at the neurocognitive level. Overall, this thesis reveals that morphosyntax, tested through SUBIRR number agreement, is especially impaired in French-speaking teens with DLD when compared to their TL peers. We discuss the findings in relation to clinical practice and highlight the importance of examining neurocognitive processes in language assessment.
Keywords: Adolescence, assessment, developmental language disorder, event-related potentials, French, morphosyntax, neurocognition, semantics, speech-language pathology, subject-verb agreement.

Table of contents

List of figures

The figures are numbered by section.

Manuscript #1

Manuscript #2

Manuscript #3

List of tables

The tables are numbered by section.

Manuscript #1

Manuscript #2

Manuscript #3

List of abbreviations

AN:	Anterior negativity
DP:	Declarative-procedural model
CONS:	Consonant-final verbs
DLD:	Developmental language disorder
EEG:	Electroencephalogram
ERP:	Event-related potentials
fLEX:	Assessment of inFlectional and LEXical processing
GSH:	Generalized slowing hypothesis
LAN:	Left anterior negativity
LIAIS:	Liaison verbs
NP:	Noun phrase
PDH:	Procedural deficit hypothesis
PÉ:	Potentiels évoqués
PP:	Prepositional phrase
SLI:	Specific language impairment
SUBIRR:	Subregular and irregular verbs
TL:	Typical language
UdeM	Université de Montréal

Acknowledgements

This (almost) 6-year project was a wonderful adventure with great challenges. In 2011, on Thursday 28 April, Phaedra Royle interviewed me to be a research assistant in her lab. After she hired me, we met by chance that same week on our bikes on Rachel Avenue and then at a concert venue. Our collaboration was meant to be. Thank you, Phaedra, for trusting me with my unconventional CV, showing me all the bases of the event-related potentials technique and for being the quickest to read, comment and revise all the grant proposals and manuscripts I have sent you over the years - you are the best PhD supervisor and mentor anyone could ask for. I thank my co-supervisor Karsten Steinhauer for his dedication to reading my manuscripts carefully and always raising issues that take time to resolve but deeply improved my work. Thanks to both of you for being such a smart and cool team and for always raising the bar for your students.

I would like to thank my husband Guillaume Loignon who takes the time to explain machine learning statistical analysis to me over breakfast, lunch and dinner. I could not have achieved so much without your wisdom and encouragement! Thanks also to Guillaume and our good friend Yves-André Bergeron, who accepted in exchange for a poutine to write more than a thousand lines of R code so that I could cross my auditory stimuli with EEG recordings, this code will have a long life.

Thanks to my father Jacques, who sparked my interest in computer science by showing me how a ZX81 microcomputer worked when I was 10 years old and to my mother Louiselle, who shared her with me her sense of humour and interest in helping others. Thanks to my sister Anne and all my friends who helped me laugh through the difficult times of my PhD project. A special thanks to my friend Ariane who suffered 4 hours under the cap while I was piloting the study.

Many thanks to all the students who worked with me. Their help was indispensable for the success of this project: Guillaume Blais, Ariane Brucher, Ann-Sophie Duquette, Marielle Julien, Khadidja Meftah, Clara Misirliyan, Cynthia Sarkis and Mélanie Weiss. A special thanks to Lisa Martignetti, my co-author on the adult manuscript, it was really fun working with you! Special thanks also go to Christophe Chénier, Lauren Fromont and Alex Herbay for their help and support.

And above all, many thanks to all the participants who contributed to this dissertation. To all the adolescents with DLD who didn't let their language impairment stop them from participating in my experiments with a smile: Thank you! I will long remember your excellent comments and questions; for example, whether I could read minds with the ERP cap, or this participant who saved a print screen of his Google Chrome's offline T-Rex game score on the testing computer desktop because it was his personal high score. Or this other participant who asked me if the ERP gel was toxic, and after my negative answer, tasted the gel and commented on the salty taste and its connection to electrical conductivity: a real young scientific mind.

Funding acknowledgements

The research projects presented in this thesis received financial support from Social Sciences and Humanities Research Council (SSHRC grant # 435-2015-1280 (PI : Phaedra Royle, Karsten Steinhauer et al "Tracking Language Acquisition and Processing in Specific Language Impairment"), and CRBLM—*COVID funding* (PI : Phaedra Royle, Karsten Steinhauer et al. for "Tracking Language Acquisition and Processing in Specific Language Impairment", 2021-2022).

Emilie Courteau was supported by a Social Sciences and Humanities Research Council's (SSHRC) doctoral fellowship and the *Fonds de recherche du Québec – Société et culture* (FQRSC)'s doctoral research scholarship. She received a graduate student stipend from the Centre for Research on Brain, Language and Music (CRBLM, Montréal), three excellence scholarships from the Biomedical Sciences Program at the University of Montreal, and numerous conference-travel awards from the CRBLM, the REPAR, Words in the World, and the University of Montreal's Faculty of Graduate Studies (FESP).

Author contributions

Manuscript #1: The first manuscript is authored by Émilie Courteau, Guillaume Loignon, Karsten Steinhauer and Phaedra Royle. ÉC, PR and KS selected the 20 tasks on which the manuscript is based. ÉC collected the data with the help of research assistants, developed the research objectives and wrote the manuscript. ÉC and GL developed the methodology and GL performed the analyses. PR and KS edited and reviewed all steps of manuscript writing.

Manuscript #2: All four authors, Émilie Courteau, Lisa Martignetti, Phaedra Royle and Karsten Steinhauer designed the experiment. EC and LM collected the data. They both performed analyses and statistics on data: LM did preliminary analyses of 16 participants and performed statistical analyses on 10 conditions with 4 ANOVAs distributed between 2 time-windows as part of her master's thesis. EC extended these preliminary analyses to the final version of the manuscript which included 28 participants, 18 conditions, and 34 ANOVAs distributed on 16 time-windows. LM wrote the first draft of the introduction, the methods section, and certain result sections as a part of her master's thesis. KS and PR oversaw all stages of data analysis. EC, PR and KS wrote and edited the final version of the paper. EC, PR and KS developed the theoretical background and issues addressed in the paper, and wrote the results and discussion of the manuscript.

Manuscript #3: The third manuscript is authored by Émilie Courteau, Phaedra Royle and Karsten Steinhauer. ÉC collected the data with the help of research assistants, developed the research objectives, performed the statistical analyses and wrote the manuscript. PR and KS oversaw all stages of data analysis and edited and reviewed all steps of manuscript writing.

1. General introduction

This dissertation is organised as follows: first, we present literature associated with the central concepts of the thesis. Second, we specify the objectives and provide an overview of the three manuscripts included in this dissertation. The subsequent sections consist of the three manuscripts, interspersed with bridges summarising findings and transitioning to following sections. Finally, we provide a general discussion including a summary of the manuscripts' results in relation to its objectives, and discuss clinical implications stemming from the thesis.

1.1. A common denominator for DLD

Finding a defining characteristic of neurodevelopmental language impairment is not an easy task, as children around the world learn different languages. An illustration of this challenging enterprise can be found in the recent change of the terminology used to describe this disorder. The label specific language impairment (SLI, e.g., Leonard, 1981), which became widely used in the 1980s (Reilly et al., 2014), defines the language-learning impairment based on two main characteristics namely that (1) cognitive difficulties are specific to language and (2) will arise in the absence of other developmental deficits, i.e., non-verbal deficits such as in the case of a global learning disability (*ibid*).

However, after decades of research, it has been argued that the term "specific" is misleading as no evidence supporting a "genetic" language impairment profile distinct from an "environmental" one has been found (Bishop et al., 2017, p. 1076). This and concerns about children's access to services motivated a consortium of 59 experts representing ten disciplines, including speech-language pathology (SLP), psychology and education, to instead propose use of the label developmental language disorder for children with language-learning deficits (DLD; Bishop et al., 2016, 2017). In the new terminology for DLD, two previous SLI characteristics have been updated and the disorder is now defined as follows: (1) functional impairments resulting from language difficulties that (2) can be accompanied by other deficits not associated with a known biomedical condition. More precisely, DLD can coexist with other neurodevelopmental disorders such as attentional deficits, but it must not be associated with conditions with a clear biomedical aetiology such as, e.g., autism spectrum disorder (Bishop et al., 2017). The DLD label has been rapidly integrated into practice by language acquisition researchers (Volkers, 2018) as well as clinical SLPs (McGregor et al., 2020), including those from Quebec[1]. Since this thesis is concerned with issues specific to language acquisition in research and clinical settings, we will follow suit and use the DLD label as defined by Bishop et al. (2017).

1.1.1. Subject-verb agreement and DLD

Although a defining characteristic has not been agreed upon for DLD, and although children with DLD exhibit different linguistic deficits across languages, cross-linguistic reviews have revealed weaknesses in phonological short-term memory and grammatical processing (comprehension or production, based on multiple languages including English, French, Italian and German: Leonard, 2014; Balilah et al., 2019). Within the domain of grammatical processing deficits, much attention has been paid to subject-verb agreement as a hallmark of DLD[2], as evidenced by numerous theories of DLD focused on this feature. Clahsen proposed the “missing agreement account” (Clahsen, 1989; Clahsen and Hansen, 1993), a theory of agreement deficits

[1] The Quebec College of Speech-Language Pathologists and Audiologists (Ordre des Orthophonistes et Audiologistes du Québec, [OOAQ]) recommends that SLPs use the DLD label (OOAQ, 2018). See Breault et al. (2019) for a review and discussion about the labels used by SLPs and other professionals in French-speaking countries, and their implementation.

[2] For simplicity, we will use the term DLD when reviewing studies that have previously used the label “specific language impairment”.

in DLD based on spontaneous speech in German-speaking children. It suggested that these children have problems establishing structural relations between two elements where one asymmetrically controls the other, as in subject-verb agreement where the subject noun "controls" the verb, establishing its number and person features. This theory was later challenged by Rice et al. (1997), who noted that, contrary to what was proposed by Clahsen (1989), the full subject-verb agreement paradigm must be available to German-speaking children with DLD: in their utterances, both correctly produced finite verbs and erroneous non-finite verbs are found. Rice et al. (1997) suggested that agreement deficits in DLD were better explained by the "extended optional infinitive" account (Rice et al., 1995). This account states that finiteness markers are omitted in DLD just as in children with TL, but for an extended period during development. Based on predictions of this account for French-speaking children with DLD, Paradis and Crago (2001) proposed to rename it the "extended optional default" account. While agreeing that children with DLD make the same errors as children with TL but for an extended period, they suggested the term "default" to account for the overuse of inappropriate finite verb stems in French-speaking children with DLD. These finite verb stems (present tense singular forms) appear to be used in French in lieu of root infinitives found in English, alongside non-finite verb forms. For a review of cross-linguistic data supporting these three accounts, see Balilah et al. (2019).

In parallel, other studies focused on the larger domain of grammatical morphology, i.e., morphosyntax, in English-speaking children with and without DLD. It was proposed that morphosyntax was a specific area of weakness when compared to lexico-semantics. Bedore and Leonard (1998) analysed spontaneous speech of children with and without DLD aged 3;7 to 5;9 (years;months), and they argued that a composite score based on their use of verb morphology discriminated better between children with and without DLD than a composite score based on their use of noun morphology, suggesting that the verb inflection subdomain of morphosyntax was impaired in DLD. In a subsequent study using similar methods and participants, Leonard et al. (1999) indicated that mastery of both verb and noun inflection (i.e., morphosyntax more generally) fell below expectations based on lexical diversity in children with DLD. Similarly, the status of morphosyntax as a hallmark of DLD was also demonstrated in older children. Conti-Ramsden et al. (2001) showed that four main markers were reliable for identifying DLD in 11-year-old children: repetition tasks of nonwords and sentences, and two tasks specifically targeting morphosyntax, i.e., third person singular and past tense verb production.

Overall, these results suggest that difficulties in subject-verb agreement is a hallmark of English-speaking preschoolers with DLD aged 5, and even older children with DLD, and that it is an area of particular importance when compared to lexico-semantics. Following these findings, more recent studies were conducted in French. French is a suitable language to study subject-verb agreement as its verb inflection system is richer than that of English and could influence the developmental trajectory in children with and without DLD. More specifically, French has multiple verb conjugation groups, as well as tense, number and person marking either marked on the verb or through pronouns and pronoun liaison, as well as "simple" (inflection morphology) and "compound" tense marking (with auxiliaries). Although multiple studies have examined the development of French morphosyntax in youth with DLD (e.g., Jakubowicz & Roulet, 2007; Royle et al., 2017), we will focus on those that have compared morphosyntactic and lexico-semantic skills.

Elin Thordardottir and Namazi (2007) analysed the spontaneous speech of French-speaking children with and without DLD, aged 3;11 on average. Both groups made very few errors on grammatical morphology, and the authors concluded that morphological deficits in French do not appear to be an area of particular weakness in children with DLD. In a subsequent study targeting 5-years-old French-speaking children with and without DLD, Elin Thordardottir et al. (2011) found that the DLD group made significantly more errors on tasks assessing comprehension and production of subject-verb agreement than their peers with typical language. Nevertheless, these tasks were not identified as useful in discriminating preschoolers with and without DLD. Rather, a combination of a receptive vocabulary task and a nonword repetition task was found to be most sensitive to the presence of a language disorder in young children. These two studies suggest that, in French DLD preschoolers, subject-verb agreement and morphosyntax in general, should not be viewed as especially impaired when compared to lexico-semantics. However, using elicitation tasks, Royle and collaborators (Royle et al., 2018; Royle and Reising, 2019; Royle and Thordardottir, 2008) have shown that in speech production, morphosyntactic agreement within the noun-phrase as well as past tense verb inflection are impaired in French-speaking children with DLD aged 4 and 6. Regarding older children with DLD, Rose and Royle (1999) evaluated morphosyntactic skills of 20 participants from families with DLD aged 9 to 46 years, and 8 controls matched on age and educational background. They analysed their spontaneous speech and tested their morphosyntactic skills using verb tense and derived word elicitation tasks, sentence comprehension, and grammaticality judgement. As participants with DLD made significantly more errors than the control group, the authors concluded that a morphological processing deficit for DLD is supported by French data. However, these studies narrowly focused on specific morphosyntactic processes and did not assess other linguistic domains, including lexico-semantics.

Overall, unlike English, French studies suggest that subject-verb agreement, and morphosyntax generally, are not areas of particular weakness in spontaneous speech-production in preschoolers with DLD when compared to lexico-semantic skills (Elin Thordardottir et al., 2011; Elin Thordardottir & Namazi, 2007). Although Royle and collaborators (1999, 2008, 2018, 2019) suggest that morphosyntax is impaired when evaluated with elicitation tasks in pre-school and older children or even adults with DLD, morphosyntax has yet to be established as an area of particular weakness in French. Furthermore, the link between morphosyntax and other language domains, including lexical-semantics, has not been explored in French-speaking children older than 5 years. In French, there is evidence that one aspect of subject-verb agreement, irregular subject-verb number agreement, is mastered only around 8 years of age for children with typical language (TL) development. This opens the door to the possibility that morphosyntactic deficits are a hallmark of French DLD only at the end of childhood and at the beginning of adolescence.

1.1.2. Subject-verb number agreement in French

Before reviewing the one study that addressed the acquisition of irregular subject-verb number agreement in French, we provide a brief presentation of the French language conjugation system. French verbs include regular, subregular and irregular agreement patterns distributed across three conjugation groups (Royle et al., 2018). Regular verbs are found only in the first conjugation group, their infinitival form ends in *-er* (like *mang-er* ‘to eat’), and they are considered to be the default form in French (Royle et al., 2012). All verbs in this group, except *aller* ‘to go’ and *s’en aller* ‘to leave’, maintain the same phonological form when in the singular or plural 3rd person across tenses (e.g., *manger* ‘to eat’: singular, *il mange* ‘he eats’ [ilmãj]; plural, *ils mangent* ‘they eat’ [ilmãj]). The second conjugation includes subregular verbs characterised by an infinitival form ending in *-ir* (like *rugir* 'to roar’). These maintain the same phonological stem throughout their conjugation and have an audible consonant-final plural form -*ssent* /-s/ in the 3rd person plural (e.g., *rugir* ‘to roar: singular, *il rugit* [ilʁyʒ**i**] ‘he roars’; plural, *ils rugissent* [ilʁyʒɪ**s**] ‘they roar’). Note that the second group also include irregular verbs, which have an infinitival form ending in -ir but do not have the -ssent /-s/ ending in the 3rd person plural (e.g., couvrir ‘to cover: singular, il couvre [ilkuvʁ] ‘he covers; plural, ils couvrent [ilkuvʁ] ‘they cover’). The third conjugation group include irregular verbs, whose stems can have various thematic vowels, undergo vowel changes, or have plural forms with final consonants other than in /-s/ (e.g., pondre ‘to lay eggs’, elle pond [ɛlpɔ̃] ‘she lays (eggs)’, elles pondent [ɛlpɔ̃d] ‘they lay (eggs)’). In the present thesis, we have used subregular and irregular verbs from the second and third conjugation groups, henceforth SUBIRR, to study subject-verb number agreement. We have combined these two groups because they both have different phonological forms in the production of singular and plural present tense, as did other authors (e.g., Franck et al., 2004, Pourquié, 2015).

Acquisition of present tense SUBIRR subject-verb number agreement by French-speaking children with TL appears to consolidate only in later childhood. Using a sentence-completion task for subject-verb number agreement supported with visual scenes, Franck et al. (2004) elicited the production of the highly frequent irregular verb *faire* ‘to do’ (*il fait* [fɛ] ‘he does’ vs. *ils font* [fɔ̃] ‘they do’). Comparing 60 TL children aged 5 to 8;5 and 8 children with DLD aged 5 to 9;4 (aged 8;8 on average, no individual data is presented), they found that TL children aged 7 were still producing 25% subject-verb agreement errors, whereas this number dropped to 5.4% by age 8;5. Adults produced errors 3.5% of the time, although the details of these results are not available (unpublished data cited in Franck et al., 2004). These results indicate that children with TL will produce subject-verb number agreement on the verb *faire* ‘to do' similarly to adults only around age 8;5. Children with DLD in this study still produced errors 20% of the time. Since this study was run on participants within a wide age range and targeted only one irregular verb, mastery of SUBIRR verbs by older French children and teenagers with and without DLD remains to be understood.

In this thesis, SUBIRR subject-verb number agreement skills in French-speaking pre-adolescents and adolescents with and without DLD were examined in relation to other morphosyntactic and lexical-semantic skills, as there is a knowledge gap in this area.

1.2. Thesis objectives

In the present thesis, we will address the knowledge gap surrounding SUBIRR subject-verb number agreement skills in French speaking (pre-)teenagers with and without DLD through three general objectives. We will (1) determine if French-speaking (pre-)teenagers with TL have acquired SUBIRR subject-verb number agreement; (2) resolve whether French-speaking (pre-)teenagers with DLD have impaired SUBIRR subject-verb number agreement skills when compared to teens with TL, and (3) establish whether SUBIRR subject-verb number agreement, and more generally morphosyntax, is a particular weakness in (pre-)teenagers with DLD when compared to other linguistic domains.

In the next sections, we will first present the two main categories of instruments available to SLPs and researchers in the field of language acquisition when evaluating language and more specifically SUBIRR subject-verb number agreement. We will then present two theories from which clear predictions about morphosyntactic and lexico-semantic processing in (pre-)teenagers with and without DLD can be derived.

1.3. Instrumentation used to assess morphosyntactic development

The linguistic theories on subject-verb agreement deficits in DLD that emerged in the 1990s (see section 1.1.1) primarily described the linguistic behaviour of children and did not question underlying brain processes. According to Chierchia (2001), linguistics collects and classifies facts about human languages. The cognitive turn in linguistics gave rise to psycholinguistics, which asks questions about the mental mechanisms responsible for these linguistic facts (*ibid*). The present thesis uses a psycholinguistic approach, where morphosyntax is assessed from a behavioural level, using linguistic tasks, and from a neurocognitive level, via event-related potential (ERP) measurements.

1.3.1. Behavioural linguistic tasks

In their recommendation on language assessment for DLD, Bishop et al. (2017, *Statement 11*) state that one should evaluate the main domains of language listed as the following: phonology, (morpho-)syntax, word finding and semantics, pragmatics/language use, discourse and verbal learning and memory. In order to provide a language profile of (pre-)adolescents with and without DLD and to fulfill the thesis' third objective of establishing whether morphosyntax is a particular weakness in (pre-)teenagers with DLD, we used 20 tasks to assess these language domains, minus pragmatics and discourse. Since this thesis focuses on morphosyntax, we were interested in language processing at the word and sentence level. This is why we did not assess participants' pragmatic and discourse skills, as these domains can be seen as higher levels of language processing (Justice et al., 2015), and these areas were considered too distant from our domains of investigation.

To evaluate SUBIRR subject-verb number agreement skills of (pre-)teenagers with and without DLD, we first used three tasks from the fLEX assessment tool (fLEX: Assessment of inFlectional and LEXical processing; Pourquié, 2015). The fLEX test is a research tool and does not provide norms, but we selected it because it assesses subject-verb number agreement with 20 SUBIRR verbs and avoids linguistic cues other than inflection for task achievement. No standardised test has been developed for French with such a wide variety of French verbs. We present this tool in manuscript #1, and in supplementary materials for the same manuscript we describe in detail the three tasks from the fLEX battery and the other tasks used to assess participants. Second, we used a task of grammaticality judgement to assess SUBIRR number agreement skills. More precisely, participants judged the acceptability of pairs consisting of grammatical spoken sentences presented along with a picture that either matched or mismatched the subject number of the sentence. This task was composed of 60 SUBIRR verbs, which included 13 of the verbs also used in fLEX. The acceptability judgement task is presented in detail in manuscript #2 and its supplementary materials, as well as in manuscript #3.

1.3.2. Neurocognitive event-related potential experiment

Imagine that two adolescents, one with TL and the other with DLD, are being assessed on their SUBIRR subject-verb number agreement skills with behavioural tasks, as we did for the present thesis. The participants will hear, for instance, *ils rugissent* [ilʁyʒɪ**s**] ‘they roar’, and will need to point to the picture that represents either one or two lions roaring. It is reasonable to imagine that both participants succeed on this task, as they are adolescents. With behavioural tasks such as this, it is impossible to know whether participants with DLD use the same language processing strategies as youth with TL. Indeed, these tasks are sensitive only to effects that can be measured with observable behavioural measures. Because the event-related potentials (ERP) technique allows us to study cognitive processes underlying performance on behavioural tasks and offers an excellent temporal resolution, it enables us to understand when and how different linguistic operations unfold over time (Steinhauer & Connolly, 2008).

Another advantage of the ERP technique is that we can compare groups on neurocognitive processes underlying multiple linguistic structures and domains. From the electroencephalogram (EEG) signal, we isolate ERP waves linked to stimuli of interest, as for instance the onset of a SUBIRR verb, e.g., rugissent [ʁyʒɪs] 'roar'. Over the years, these wave patterns, called "ERP components", have been consistently observed in different languages for similar linguistic structures. These established ERP components are assumed to reflect specific neurocognitive processes (Steinhauer & Connolly, 2008). Furthermore, relatively different ERP patterns have been associated with lexico-semantics and morphosyntactic processing (ibid).

By comparing ERP components found in adolescents with and without DLD, we can determine if they are using similar or different neurocognitive mechanisms to process agreement, for example. Similarly, we can establish whether adolescents with TL process information, such as SUBIRR subject-verb number agreement, in a mature, and automatic manner by comparing their ERPs with those of adults. We can also compare ERP patterns between linguistic domains and determine if DLD participants resemble their peers with TL more within morphosyntax or lexico-semantic domains. For a literature review of ERP components underlying lexico-semantic and morphosyntactic processing in adults, see manuscript #2, and for a comprehensive review of ERP in the use of the assessment of morphosyntactic development in children with and without DLD, see manuscript #3.

One more benefit of the ERP technique is that it is suitable for developing an innovative paradigm with a high ecological validity in th e study of language. ERP studies usually present ungrammatical sentences to their participants. However, children and adolescents with and without DLD usually encounter grammatical sentences in their everyday life. To use a more ecological approach to language evaluation than presenting ungrammatical sentences, we developed an ERP experiment that used only grammatical sentences. Auditory-visual subject number mismatches were created by combining an inappropriate number of visually presented subjects concurrently with morphosyntactically correct number cues in the auditory stimuli. We expected that this more naturalistic experimental paradigm would yield results closer to everyday language use than the artificial context of ungrammatical sentences (Kandylaki and Bornkessel-Schlesewsky, 2019). Indeed, the combination of pictures and speech resembles other common activities such as shared picture-book reading or watching movies or videos, where a picture, animation, or video is presented along with a narration. In these cases, people being read or spoken to might have expectations about what they will hear, and notice any incongruencies, as our participants were expected to do during this experiment. This innovative paradigm is presented in detail in manuscript #2 and in the supplementary materials accompanying it.

1.4. Two theories of language processing in DLD

For a psychological theory to be tested or falsified, it must be formulated in such a way that we can derive clear predictions from it (Eronen & Romeijn, 2020). As presented above, many theories from which clear predictions could be derived were proposed in the 1990s to explain impaired subject-verb agreement in DLD (see section 1.1.1). However, the scope of these theories generally extended only to subject-verb agreement abilities within the DLD, or more generally only to morphosyntax. Within the third objective of this thesis, we are interested in comparing morphosyntax with other linguistic domains such as lexico-semantics in order to determine whether morphosyntax is a particular weakness in French DLD. To do this, we needed theories with a broader scope. We assessed our participants' morphosyntactic abilities within the framework of two theories spanning behavioural and neurocognitive domains that allowed us to make clear predictions about morphosyntactic and lexical-semantic processing in adolescents with and without DLD. We selected two theories because it makes more sense for us to interpret our results in a comparative approach rather than to confirm or refute one theory by itself (see Rosenberg, 2005, p. 119, for a discussion on hypothesis testing in science as a comparative enterprise).

1.4.1. The procedural deficit hypothesis

The procedural deficit hypothesis (PDH; Ullman & Pierpont, 2005; Ullman et al., 2020) suggests that, in DLD, morphosyntax will be impaired in contrast to relative strengths in lexico-semantics. According to the PDH the procedural memory system is responsible for morphosyntactic processing. In contrast, the PDH assumes that the declarative memory system, underlying lexico-semantic processing, remains relatively normal in DLD and can compensate for the deficits in the procedural system. This hypothesis is based on the declarative-procedural model (DP; Ullman, 2004, 2020), which was developed with the aim of identifying cognitive domains that share commonalities with language and whose underlying systems would be promising candidates for those that support language (Ullman, 2004). The model proposes that the declarative memory system is responsible for the use and knowledge of facts and events, and thus of memorised information as that within the mental lexicon, i.e., lexico-semantics. The procedural memory system underpins, among others, the learning and the processing of habits, perceptual sequences as in statistical learning and perceptual-motor skills. Within language, this system is largely responsible for (morpho-)syntax, speech-sound representation and production.

The PDH posits that abnormalities in persons with DLD's brain structures supporting the procedural memory system explain this dichotomy between language domains as well as many of the behavioural characteristics in children with DLD (e.g., motor impairments, temporal processing deficits, Ullman & Pierpont, 2005, or execution functions such as inhibition and working memory, Ullman et al., 2020).

Within morphosyntax, at the word-form level, the procedural system supports rule-governed regular morphology processing (inflection, and possibly derivation) such as English third person verb number agreement (e.g., *she sing-s*) and, in French, plural verb number agreement instantiated through liaison of the plural pronoun to the following verb with a vowel onset (e.g., *elle achète* [ɛlaʃɛt] ‘she buys’ vs. *elles achètent* [ɛl**z**aʃɛt] ‘they buy’, see manuscript #2). Unpredictable word forms as in English *she sang*, or in the French irregular verb *ils pondent* [ilpɔ̃**d**] (‘they lay [eggs]’) are argued to be stored in the lexicon, which is supported by the declarative memory system. At the sentence-level, the procedural system underlies agreement–such as number or tense–between sentence constituents (Steinhauer & Ullman, 2002). In children with TL, rule-governed morphosyntactic information either at the word-form or sentence level is assumed to first be at least partially supported by the declarative system. It is gradually taken over by the procedural system, which leads to automatisation (Ullman, 2020). In the case of DLD, a procedural memory deficit would imply that rule-governed morphosyntactic information will not be processed by this system; instead, this type of information will be stored and processed by the declarative memory system as a compensatory strategy (Ullman et al., 2020; Ullman & Pierpont, 2005).

Note that within lexico-semantics, the procedural memory is expected to underpin lexical retrieval (Ullman and Pierpont, 2005). It is thus expected that tasks involving lexical retrieval are likely to be more affected in DLD than receptive tasks such as word comprehension or word recognition. In addition, if lexical retrieval tasks do not require rapid responses, the PDH predicts that improved and possibly normal performance should be observed in participants with DLD (*ibid*, p.418).

1.4.2. The generalized slowing hypothesis

In contrast to the PDH, the generalized slowing hypothesis (GSH; Kail, 1994) posits that all language domains are impaired in DLD. Kail (1994) reanalysed 22 experimental conditions testing linguistic domains and verbal memory and compared reaction times of children with and without DLD. As his results showed that the DLD group were consistently slower than the TL one, Kail (1994) suggested that the general cognitive component of processing speed is impaired in DLD, thus explaining slower reaction times in children with DLD when compared to those with TL. The GSH proposes that children with DLD use the same processes as TL children to execute a given task; however, the time needed to execute each process will be multiplied by a common coefficient resulting in a generalized slowing effect. A slower processing speed will result in processing limitations because, as the information is not processed fast enough, it will be vulnerable to degradation or interference from other incoming information (Kail & Salthouse, 1994). Originally, Kail (1994) proposed that a slower processing speed in DLD would affect linguistic and non-linguistic domains in a similar way. However, Kail updated his view in a study with Leonard et al. (2007) using multiple tasks administered to 14-year-olds with and without DLD and confirmatory factor analyses. Their results suggest that linguistic processing speed can be differentiated from non-linguistic processing, and therefore a common slowing factor cannot apply to all cognitive domains. However, a similar delay across all language domains is expected. These authors concluded that processing limitations can compound children's language difficulties and, perhaps, be a primary cause for the impairment.

1.5. Overview of the thesis manuscripts

In the first manuscript, we address all three thesis objectives at the behavioural level. We investigate SUBIRR subject-verb number agreement skills of French-speaking (pre-)teenagers with and without DLD in relation to their other morphosyntactic abilities in as well as other linguistic domains. To do so, we administered 20 subtasks that assessed linguistic, metalinguistic and working memory skills in two groups of participants: 17 adolescents clinically identified as having DLD (age M = 14;09) and 20 (pre-)teens with TL (age M = 12;21). In-depth details on participants and all behavioural subtasks used in the present thesis can be found in the first manuscript as well as its supplementary materials. In order to identify SUBIRR subject-verb number agreement and morphosyntactic deficits as a potential clinical marker of French DLD, we used a methodology typical of clinical studies for diagnostic test accuracy, including measures of sensitivity, specificity as well as positive and negative likelihood ratios. We also included new robust statistical methods that are less affected by outliers (e.g., permutations and regularisation).

In the second manuscript, we investigated neurocognitive processes underlying determiner-noun and subject-verb number agreement in French-speaking adults. In order to achieve the first objective of this thesis, namely to determine if French-speaking (pre-)teenagers with TL have acquired SUBIRR subject-verb number agreement, we needed to know what to expect from adults. Twenty-eight French-speaking adults participated in the experiment. We developed, as previously described, an innovative experimental paradigm using only grammatical sentences to assess morphosyntactic and lexico-semantic verb processing. Lexico-semantic mismatches were created by presenting a verb that did not match the depicted action. Number agreement mismatches were created by mismatching the number of visually presented subjects and morphosyntactic number cues in the auditory stimuli, either in the noun phrase at sentence onset, or on the verb downstream in the sentence. Subject-verb number agreement was marked on plural verb forms either through regular liaison or with SUBIRR verb-final consonants. After each sentence, participants made an acceptability judgement on whether the sentence matched the depicted picture. We recorded ERPs and expected to elicit classic ERPs for agreement errors although we used only grammatical sentences.

In the third and last manuscript, we addressed all three thesis objectives at the behavioural level, using acceptability judgements, and at neurocognitive level, using the ERP experiment presented in manuscript #2. We articulated predictions based on the previous literature on morphosyntactic development in ERP studies with children as well as on the two chosen theories where clear neurocognitive predictions about morphosyntactic and lexico-semantic processing in DLD could be derived. The PDH proposes that morphosyntactic impairments are related to a procedural memory deficit for grammatical processing, with a preserved declarative memory for lexico-semantic abilities. The PDH also allows us to contrast highly-regular and SUBIRR number agreement, as the former is expected to be related to the procedural system, whereas the latter would be associated with the declarative system. The GSH proposes that (pre-)teenagers with and without DLD will use similar processing neurocognitive processes for both linguistic domains, but they will be characterised by limited processing speed for the DLD group, and thus incur timing delays for ERP components.

1.6. References

Balilah, A., Rafat, Y., & Archibald, L. (2019). Domain-Specific and Domain-General Processing Accounts in Children with Specific Language Impairment (SLI): Contribution of Cross-Linguistic Evidence. In S. Hidri (Ed.), *English Language Teaching Research in the Middle East and North Africa: Multiple Perspectives* (pp. 383–407). Springer International Publishing. https://doi.org/10.1007/978-3-319-98533-6_18

Bedore, L. M., & Leonard, L. B. (1998). Specific Language Impairment and Grammatical Morphology. *Journal of Speech, Language, and Hearing Research*, *41*(5), 1185–1192. https://doi.org/10.1044/jslhr.4105.1185

Bishop, D. V. M., Snowling, M. J., Thompson, P. A., & Greenhalgh, T. (2016). CATALISE: A Multinational and Multidisciplinary Delphi Consensus Study. Identifying Language Impairments in Children. *PLoS ONE*, *11*(7). https://doi.org/10.1371/journal.pone.0158753

Bishop, D. V. M., Snowling, M. J., Thompson, P. A., & Greenhalgh, T. (2017). Phase 2 of CATALISE: A multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *Journal of Child Psychology and Psychiatry*, *58*(10), 1068–1080. https://doi.org/10.1111/jcpp.12721

Breault, C., Béliveau, M.-J., Labelle, F., Valade, F., & Trudeau, N. (2019). Le trouble développemental du langage (TDL): Mise à jour interdisciplinaire. *Neuropsychologie Clinique et Appliquée*, *3*(automne 2019)), 64–81. https://doi.org/10.46278/j.ncacn.20190717

Chierchia, G. (2001). Linguistics and Language. In R. A. Wilson & F. C. Keil (Eds.), *The MIT encyclopedia of the cognitive sciences* (pp. 92–111). MIT press.

Clahsen, H. (1989). The grammatical characterization of developmental dysphasia. *Linguistics*, *27*, 897–920. https://doi.org/doi.org/10.1515/ling.1989.27.5.897

Clahsen, H., & Hansen, D. (1993). *The missing agreement account of specific language impairment: Evidence from therapy experiments* [Monograph]. http://repository.essex.ac.uk/242/

Conti-Ramsden, G., Botting, N., & Faragher, B. (2001). Psycholinguistic markers for specific language impairment (SLI). *Journal of Child Psychology and Psychiatry*, *42*(6), 741–748. https://doi.org/10.1111/1469-7610.00770

Elin Thordardottir, Kehayia, E., Mazer, B., Lessard, N., Majnemer, A., Sutton, A., Trudeau, N., & Chilingaryan, G. (2011). Sensitivity and specificity of French language and processing measures for the identification of primary language impairment at age 5. *Journal of Speech, Language, and Hearing Research*, *54*(2), 580–597. https://doi.org/10.1044/1092-4388(2010/09-0196)

Elin Thordardottir, & Namazi, M. (2007). Specific language impairment in French-speaking children: Beyond grammatical morphology. *Journal of Speech, Language, and Hearing Research*, *50*(3), 698–715. https://doi.org/1092-4388/07/5003-0698

Eronen, M. I., & Romeijn, J.-W. (2020). Philosophy of science and the formalization of psychological theory. *Theory & Psychology*, *30*(6), 786–799. https://doi.org/10.1177/0959354320969876

Jakubowicz, C., & Roulet, L. (2007). Narrow syntax or interface deficit? Gender agreement in French SLI. In J. Liceras, H. Zobl, & H. Goodluck (Eds.), *The Role of Formal Features in Second Language Acquisition* (Routledge, pp. 184–225). Routledge. https://doi.org/10.4324/9781315085340-7

Justice, L. M., Lomax, R., O'Connell, A., Pentimonti, J., Petrill, S. A., Piasta, S. B., Gray, S., Restrepo, M. A., Cain, K., & Catts, H. (2015). The dimensionality of language ability in young children. *Child Development*, *86*(6), 1948–1965. https://doi.org/10.1111/cdev.12450

Kail, R. (1994). A Method for Studying the Generalized Slowing Hypothesis in Children With Specific Language Impairment. *Journal of Speech, Language, and Hearing Research*, *37*(2), 418–421. https://doi.org/10.1044/jshr.3702.418

Kail, R., & Salthouse, T. A. (1994). Processing speed as a mental capacity. *Acta Psychologica*, *86*(2), 199–225. https://doi.org/10.1016/0001-6918(94)90003-5

Kandylaki, K. D., & Bornkessel-Schlesewsky, I. (2019). From story comprehension to the neurobiology of language. *Language, Cognition and Neuroscience*, *34*(4), 405–410. https://doi.org/10.1080/23273798.2019.1584679

Leonard, L. B. (1981). Facilitating linguistic skills in children with specific language impairment. *Applied Psycholinguistics*, *2*(2), 89–118. https://doi.org/10.1017/S0142716400000886

Leonard, L. B. (2014). Specific language impairment across languages. *Child Development Perspectives*, *8*(1), 1–5. https://doi.org/10.1111/cdep.12053

Leonard, L. B., Miller, C., & Gerber, E. (1999). Grammatical morphology and the lexicon in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, *42*(3), 678–689. https://doi.org/10.1044/jslhr.4203.678

Leonard, L. B., Weismer, S. E., Miller, C. A., Francis, D. J., Tomblin, J. B., & Kail, R. (2007). Speed of processing, working memory, and language impairment in children. *Journal of*

Speech, Language, and Hearing Research, *50*(2), 408–428. https://doi.org/10.1044/1092-4388(2007/029)

OOAQ. (2018). *Le trouble développemental du langage Pour les professionnels de la santé et de l'éducation* (p. 6) [Dépliant pour les professionnels]. http://www.ooaq.qc.ca/actualites/OOAQ_deIpliant_professionnels_pour%20le%20web_v007.pdf

Paradis, J., & Crago, M. (2001). The Morphosyntax of Specific Language Impairment in French: An Extended Optional Default Account. *Language Acquisition*, *9*(4), 269–300. https://doi.org/10.1207/S15327817LA0904_01

Pourquié, M. (2015). fLEX: Multilingual assessment of inFlectional and LEXical processing. *Software, Intellectual Property 2016-01*, *88*.

Reilly, S., Tomblin, B., Law, J., McKean, C., Mensah, F. K., Morgan, A., Goldfeld, S., Nicholson, J. M., & Wake, M. (2014). Specific language impairment: A convenient label for whom? *International Journal of Language & Communication Disorders*, *49*(4), 416–451. https://doi.org/10.1111/1460-6984.12102

Rice, M. L., Noll, K. R., & Grimm, H. (1997). An Extended Optional Infinitive Stage in German-Speaking Children With Specific Language Impairment. *Language Acquisition*, *6*(4), 255–295. https://doi.org/10.1207/s15327817la0604_1

Rice, M. L., Wexler, K., & Cleave, P. L. (1995). Specific Language Impairment as a Period of Extended Optional Infinitive. *Journal of Speech, Language, and Hearing Research*, *38*(4), 850–863. https://doi.org/10.1044/jshr.3804.850

Rosenberg, A. (2005). *Philosophy of Science: A Contemporary Introduction* (Routledge). Psychology Press.

Royle, P., Beritognolo, G., & Bergeron, E. (2012). Regularity, sub-regularity and irregularity in French acquisition. In J. van der Auwera, T. Stolz, A. Urdze, & H. Otsuka (Eds.), *Irregularity in Morphology (and beyond)* (Vol. 11, pp. 227–250). Akademie Verlag.

Royle, P., & Reising, L. (2019). Elicited and Spontaneous Determiner Phrase Production in French-Speaking Children With Developmental Language Disorder. *Canadian Journal of Speech-Language Pathology & Audiology*, *43*(3). https://cjslpa.ca/files/2019_CJSLPA_Vol_43/No_3/CJSLPA_Vol_43_No_3_2019_MS_1134.pdf

Royle, P., St-Denis, A., Mazzocca, P., & Marquis, A. (2017). Insensitivity to verb conjugation patterns in French children with SLI. *Clinical Linguistics & Phonetics*, 1–20. https://doi.org/10.1080/02699206.2017.1328706

Royle, P., St-Denis, A., Mazzocca, P., & Marquis, A. (2018). Insensitivity to verb conjugation patterns in French children with SLI. *Clinical Linguistics & Phonetics*, *32*(2), 128–147. https://doi.org/10.1080/02699206.2017.1328706

Royle, P., & Thordardottir, E. T. (2008). Elicitation of the passé composé in French preschoolers with and without specific language impairment. *Applied Psycholinguistics*, *29*(3), 341–365. https://doi.org/10.1017/S0142716408080168

Steinhauer, K., & Connolly, J. F. (2008). Event-related potentials in the study of language. In B. Stemmer (Ed.), *Handbook of the neuroscience of language* (pp. 91–104). Elsevier Ltd.

Steinhauer, K., & Ullman, M. T. (2002). Consecutive ERP effects of morpho-phonology and morpho-syntax. *Brain and Language*, *83*, 62–65.

Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition*, *92*(1), 231–270. https://doi.org/10.1016/j.cognition.2003.10.008

Ullman, M. T. (2020). The Declarative/Procedural Model: A Neurobiologically-Motivated Theory of First and Second Language. In B. VanPatten, G. D. Keating, & S. Wulff (Eds.), *Theories in Second Language Acquisition: An Introduction* (3rd ed., pp. 128–161). Routledge.

Ullman, M. T., Earle, F. S., Walenski, M., & Janacsek, K. (2020). The Neurocognition of Developmental Disorders of Language. *Annual Review of Psychology*, *71*(1), null. https://doi.org/10.1146/annurev-psych-122216-011555

Ullman, M. T., & Pierpont, E. I. (2005). Specific language impairment is not specific to language: The procedural deficit hypothesis. *Cortex*, *41*(3), 399–433. https://doi.org/10.1016/S0010-9452(08)70276-4

Volkers, N. (2018). Diverging Views on Language Disorders. *The ASHA Leader*, *23*(12). https://doi.org/10.1044/leader.FTR1.23122018.44

2. Manuscript 1

Identifying Linguistic Markers of French-speaking Teenagers with Developmental Language Disorder: Which Tasks Matter?

Émilie Courteau[1,2], Guillaume Loignon[3], Karsten Steinhauer [2,4], Phaedra Royle [1,2]

[1]School of Speech Language Pathology and Audiology, Faculty of Medicine, University of Montreal, Montréal, Quebec, Canada

[2]Centre for Research on Brain, Language and Music (CRBLM), Montreal, Quebec, Canada

[3]Educationnal sciences, Faculty of Education, UQO, St-Jerome, Quebec, Canada

[4]School of Communication Sciences and Disorders, Faculty of Medicine, McGill University, Montréal, Quebec, Canada

This manuscript has been submitted to the Journal of Speech, Language, and Hearing Research for publication and is currently under review

Abstract

Purpose: Unlike in English, it is not clear whether subject-verb number agreement deficits can be viewed as a marker of developmental language disorder (DLD) in French-speaking children, as agreement may only be consolidated throughout the school-age years. Previous studies of French DLD markers targeted preschool children. The present research aimed to identify reliable tasks discriminating French-speaking adolescents from their peers with typical language (TL), and to assess which linguistic subdomain(s) represent areas of particular weakness in DLD.

Methods: We administered 20 subtasks that assessed linguistic, metalinguistic and working memory skills two groups: 17 adolescents clinically identified as having DLD (M = 14;09 [years;months]) and 20 (pre-)teens with TL (M = 12;21). Using robust statistics that are less affected by outliers, we selected the most discriminating subtasks between our groups, calculated their optimal cut-off score, and derived diagnostic accuracy statistics. We combined these subtasks in a multivariable model to identify which subtasks contributed the most to the identification of DLD.

Results: Seven subtasks were selected as discriminating between our groups, and three showed outstanding diagnostic accuracy: Recalling Sentences, a multi-word task assessing lexico-semantics skills, and an irregular verb number-agreement production task. When combined, we found that the latter, which assessed morphosyntactic skills, contributed the most to our multivariable model.

Conclusion: This study provides evidence that the most relevant markers to identify DLD in French teenagers are tasks assessing lexico-semantics and morphosyntactic domains, and that morphosyntax should be considered an important area of weakness in French-speaking teenagers with DLD.

Keywords: Adolescents, Assessment, Language disorders, Morphology, Speech-language-pathology

2.1. Introduction

The diagnostic criteria for children with language disorder in the DSM-V (American Psychiatric Association, 2013) include early onset of symptoms and persistent difficulties in the acquisition of language caused by comprehension or production deficits. These are characterized by a reduced vocabulary, limited sentence structures, and discourse impairments. Those language deficits are not the result of sensory, motor impairments, or global delay, and will result in functional limitations in many areas, including social participation and academic achievement. The most recent developmental language disorder (DLD) label suggested by Bishop et al. (2017) aligns with the DSM-V definition, and provides additional guidelines for clinicians: e.g., the first step toward a DLD diagnostic should be to establish functional impairments. Both sources converge in saying that language disorders diagnosed at the age of 4 or 5 years usually persist into adulthood. The DSM-V specifies that although the language deficit will persist, the specific profile of language strengths and weaknesses is likely to change over the child's development. How the language profile will evolve in adolescents with DLD remains unclear. Note that when we review studies that use the previous common label specific language impairment, we will translate it to DLD for the sake of clarity, even though we are aware that, for theoretical reasons, the diagnostic criteria for these labels are not interchangeable.

The vast majority of studies on DLD target language impairments has focused on preschool or young children while evidence on language outcomes for teenagers is limited (Haebig et al., 2017). The profile of French-speaking DLD teenagers is even less studied. This is a missed research opportunity since French offers the possibility to study the developmental trajectory of a phenomenon that is well known in English DLD but remains unclear in French i.e., the acquisition of subject-verb number agreement and its potential as a clinical marker of

DLD. Subject-verb agreement is known to be a significant area of weakness in English-speaking children with DLD as early as 3;1 to 6;11 ([years;months]; Leonard et al., 1999), and tasks assessing this ability have proven to be useful in identifying the disorder in older children as well (Conti-Ramsden et al., 2001). This appears to be different for French-speaking children with DLD. Morphosyntactic errors have not been identified as a salient feature of preschoolers with DLD's spontaneous speech (Elin Thordardottir and Namazi, 2007), but on tasks assessing comprehension and production of subject-verb agreement they make more errors than their peers with typical language ([TL]; Elin Thordardottir et al., 2011). However, these tasks were not identified as useful in discriminating preschoolers with and without DLD. Rather, a combination of a receptive vocabulary task and a nonword repetition task were more sensitive to the disorder in young children (*ibid*, see details below). Moreover, on the basis of an elicited production task of the irregular plural verb faire 'to do', there is evidence that subject-verb agreement is not automatized until age 8;5 for TL children (Franck et al., 2004). Since French children with TL are likely to master irregular subject-verb number agreement during school-age years, one could expect it to be a hallmark of French DLD only in (pre-)adolescence, but this has yet to be studied. Given this and the fact that only a few standardized tests are available to the French-speaking population of Quebec, including many with norms that don't meet psychometric criteria (Bouchard et al., 2009), there is an urgent need for studies that evaluate the best markers for the identification of French DLD teenagers. These markers would be useful not only for understanding the development of DLD in adolescence, but also for identifying which skills to focus on in speech-language pathology (SLP) therapy interventions.

This study aims to describe the linguistic impairments of French DLD teenagers from Quebec, Canada. Through the assessment of participants with 20 tasks often used in clinical and research settings, we intend to evaluate which ones best discriminate between pre-teens and teens with and without DLD. Our second objective is to identify which linguistic subdomains are weaknesses in French DLD adolescents, based on the underlying constructs of these tasks.

2.1.1. Language development from the dimensionality perspective

Language research on DLD and assessment tasks used by SLPs usually assume that language has multiple components i.e., it is multidimensional (Lonigan & Milburn, 2017). Indeed, many theoretical accounts have targeted specific linguistic subdomains as areas of weakness in DLD, assuming by the same token that these are up to a certain extent dissociable. For instance, the procedural deficit hypothesis (Ullman & Pierpont, 2005) proposes that aspects of (morpho-)syntax and phonology are impaired due to a deficit of the procedural memory system, whereas aspects of lexico-semantics ability linked to a typically-developing declarative memory system are preserved. In clinical settings, Tomblin and Zhang (2006) give a classic example of how language is assessed through commercially-available test batteries: domains such as grammar and vocabulary will be evaluated by different tasks in the receptive and expressive modalities, assuming that subdomains can be impaired or preserved in an individual (e.g., preserved receptive syntax versus impaired expressive vocabulary). Recently, challenging the assumption of language's multidimensionality, the development of language skills of children has been studied through linguistic assessment tasks with confirmatory factor analyses, which allow researchers to confirm if the studied constructs are distinct, and to validate if they have empirical foundations (*ibid*). In children with and without DLD, there is evidence that the multidimensionality of language increases with children's age. Looking at language's dimensions, Justice et al. (2015) tested 915 children from prekindergarten to third grade and found that only starting at grade 3, vocabulary, grammar and discourse can be considered as different dimensions. Tomblin and Zhang (2006) conducted a longitudinal cohort study of 1929 children with and without DLD with four testing points: from kindergarten through second, fourth, and eighth grades. They showed that it is valid to consider vocabulary, assessed with word-level tasks, and grammar, reflected by sentence-level tasks, as two dimensions starting in second grade, and that this multidimensionality increases with age. Lonigan and Milburn (2017) also found that vocabulary and syntax were two dimensions, but nonetheless shared a lot of variance, by assessing 1895 children with and without language impairments from pre-kindergarten to 5th grade. Both these latter studies failed to find evidence supporting the idea that comprehension and production language skills were two different dimensions. In brief, these studies agree on a multidimensional perspective of language at least in older school-aged children with and without DLD, offering support for the multidimensionality of adolescents' language abilities. Considering this, we will now review what dimensions are likely to be impaired in teenagers with DLD.

2.1.2. Linguistic impairments in French-speaking adolescents with DLD

Morphosyntactic impairments in DLD children had been found in many languages and was modelled in many theories (for reviews see Leonard, 2014 and Pourquié et al, in revision) but research was generally centred on preschool or young school-aged children. Only a few studies have detailed the morphosyntactic skills of older French-speaking children with DLD. Looking at older children allows us to see which morphosyntactic structures have been acquired and what is still impaired despite many years of practice. Tuller et al. (2012) studied spontaneous utterances by a group of 11 to 16 years-old to examine their use of syntactically complex sentences. They showed that compared to three groups of younger TL children aged 6, 8, and 11 years-old, teenagers with DLD produced significantly more ungrammatical utterances: 15.5% for the DLD group compared to under 5% for all TL groups, but crucially they also avoided using complex sentences. Less frequent use of complex morphosyntactic structures in spontaneous speech was also observed in a group of French-speaking DLD (pre-)teens aged 11;6 on average, when compared to a control group matched on morphosyntactic comprehension, aged 7;8 on average (Prigent et al., 2015). These studies suggest that avoidance of complex structures is a characteristic of French-speaking children with DLD: this underscores the importance of using elicited production tasks to assess the upper limit of these groups' language abilities. Rose & Royle, (1999) used a sentence completion task to elicit the production of 12 verbs in the present or past tense by 20 participants from families with DLD aged 9 to 46 years. They found deficits in tense production in the DLD group when compared to 8 controls matched on age and educational background. Using a completion task for subject-verb number agreement based on pictures, Franck et al. (2004) elicited the production of the irregular verb faire 'to do' (il fait [fɛ]– ils font [fɔ̃], 'he does –they do'). Based on 60 TL children aged 5 to 8;5 and 8 children with DLD aged 5 to 9;4, they found that TL children aged 7 were still producing 25% subject-verb agreement errors, whereas it dropped to 5.4% by age 8;5. Participants with DLD (aged 8;8 on average, no individual data is presented), still produced errors 20% of the time. Since these two studies were based on DLD participants within a wide age range and targeted only a small number of irregular verbs, mastery of irregular verbs by older French children and teenagers with and without DLD remains to be described.

Lexico-semantic skills, i.e., vocabulary, has not usually been thought to be a marker for DLD in school-aged children, at least in English. For instance, although Conti-Ramsden et al. (2001) administered vocabulary tasks to their participants, these were not considered potential positive psycholinguistic markers of DLD, and thus were not included in their analyses of diagnostic accuracy. However, McGregor et al. (2013) assessed 177 DLD and 325 TL children and teenagers in grades 2, 4, 8, and 10 on their vocabulary's breadth, through the number of words defined correctly, and vocabulary's depth, measured as the quantity of correct information in each definition. DLD participants showed deficits on both measures throughout all age groups. Impairments were also found on receptive vocabulary. Using the Peabody Picture Vocabulary Test (PPVT), Rice and Hoffman (2015) tested DLD and TL children, teens and young adults in a longitudinal study from ages 2;6 to 21 years, and found lower performance for participants with DLD across the duration of the study. Targeting French-speaking children, Elin Thordardottir et al. (2011) found that the EVIP (the French PPVT, Dunn et al., 1993) had a relatively high level of sensitivity and specificity for identifying 5-year-old children with DLD.

We have reviewed aspects of morphosyntax and lexico-semantics that appear to be impaired in older children and teenagers with DLD. Van Kleeck (1982) suggests that these are considered to reflect primary linguistic skills because they instantiate understanding and production of language, and can be viewed as qualitatively different from metalinguistic skills, which refers to "the ability to reflect consciously upon the nature and properties of language" (*ibid*, p. 237). It had been shown through grammaticality judgment tasks of orally-presented sentences that the metalinguistic skills of English-speaking adolescents with DLD are impaired. Miller et al. (2008) showed that English-speaking 16-year-olds with DLD were significantly less successful at identifying sentences containing subject-verb agreement omission and commission errors, compared to their TL peers matched on age. Similar findings were reported by Noonan et al. (2014) for 8 years-old with DLD and by Haebig et al. (2017) in 15-years-olds with DLD. However, the latter authors found similar performance for their DLD and TL groups on lexical-semantic errors. Metalinguistic skills of French-speaking children with DLD also appear to be poor. In a case study of a French-speaking child with DLD aged 8, Poulin et al. (2015) found that, in an oral task with visual support, his ability to identify gender agreement errors on adjectives, but not semantic errors, was impaired in comparison to age-matched TL children , but not to younger ones (aged 6). Maillart and Schelstraete (2005) observed reduced ability to detect sentences containing agreement or tense marking errors in a group of 9-year-old DLD children compared to a TL group matched on receptive grammatical skills, aged 5;4. However, they performed similarly on syntactic word order errors. Rose and Royle (1999) found that 20 (pre-)teens and adults (aged 9 to 46, from families with DLD) performed worse on the identification of determiner, preposition, verb tense, number agreement or argument-structure (e.g., missing objects) errors in sentences than a TL group. Using a lexical decision task in which participants had to identify if the word heard was a pseudoword or not, Quémart and Maillart (2016) showed that metalinguistic skills associated with the phonological lexicon of 10-year-old child with DLD were impaired when compared to TL children matched on aged or receptive vocabulary.

Oral metalinguistic judgment tasks require that participants hold auditory input in memory while they process information (Noonan et al., 2014) so it's no surprise that impairment in phonological working memory had been found in teenagers with DLD. Ebbels et al. (2012) evaluated nonword repetition in a group of 15 English-speaking DLD teenagers aged 13, and two control groups matched on chronological age or language level. They found lower performance only in half of the DLD group as compared to the two control groups. However, these results should be considered with caution because the DLD group was split in two based on a visual inspection of their nonword repetition scores prior to group comparisons in the statistical analyses. Using forward and backward digit span tasks, Arslan et al. (2020) found impaired phonological working memory skills in 2 groups of French-speaking DLD children aged 7–11 and teenagers aged 12–18 when compared to age-matched control groups. Interestingly, there wasn't any difference between the teen groups on visuospatial working memory skills assessed through the forward and backward Corsi Blocks test (1972), but the younger DLD group showed significantly lower performance than their aged-matched TL peers on the backward Corsi Blocks span, suggesting that these skills can normalize with age. Overall, these studies reveal that teenagers with DLD are likely to experience linguistic deficits expressed by impairments in morphosyntax, lexico-semantics, metalinguistics and phonological working memory. The assessment of these deficits is a challenge for clinicians in Quebec given the limited number of standardized tools available in French, a situation we will describe in the next section.

2.1.3. Language assessment of adolescents with DLD

When assessing the language of younger children that are suspected of having language impairment, one of the key outcomes is to establish a diagnostic of DLD based on functional and language impairments, as proposed by Bishop et al. (2017). Since adolescents with DLD already have been given a diagnosis, language evaluation will have other main goals. In clinical practice, the results of these language evaluations are often used to determine which language areas or structures SLP therapy should focus on. Another objective of these evaluations, at least in Quebec, is to establish language disorder severity. Even if it is not recommended by Quebec and international standards (Tessier & Valade, 2017; Bishop et al., 2017), low scores on standardized tests to confirm a disorder's severity is still a widespread criterion used in Quebec to access services (Breault et al., 2019) and is therefore a common practice among Quebec SLPs. A challenge is that only few standardized tests are available for the Quebec-French teenager population: in their review of oral language tests, Monetta et al. (2016) listed five that could be used with an adolescent population, and only one was normed and validated for the French Quebec's population, i.e., CELF-4CDN-F (Secord et al., 2009). As a result, SLPs from Quebec have the option to use adapted (or even less ideal, translated) standardized English tests or French tests standardised in France. These tests are rarely based on appropriate cultural and linguistic norms for the Quebec population (Bouchard et al., 2009). In addition to clear lexical differences, grammatical ones might also emerge. For instance, Courteau et al. (2019) showed that adult speakers of Quebec French did not systematically process incorrect omission of subject-verb plural liaison. This is not surprising since the plural feminine third person pronoun elles [ɛl/z] 'they.fem' is rarely used and often replaced (neutralized) by the masculine pronoun ils [ɪl/z] 'they.masc' in spontaneous speech (e.g., Bourget, 1987). Tasks based on different language varieties might thus under-evaluate linguistic abilities in Quebec-French speakers.

Furthermore, many tests are not based on appropriate norms that meet psychometric criteria for the target population (Bouchard et al., 2009). Indeed, Elin Thordardottir et al. (2011) showed that among 78 monolingual speakers of Quebec French, group means scores were one standard deviation higher than the published norms of the French version of the PPVT (EVIP, Dunn et al., 1993). This could be attributed to the fact that the published norms were based on pan-Canadian francophones that included monolingual but also bilingual French speakers. Thisleads to the underestimation of language difficulties of monolingual Quebec French-speaking children (Godard & Labelle, 1995, cited in Elin Thordardottir et al., 2011). Considering that 82% of the 8 million inhabitants in Quebec have French as their first spoken language (Statistics Canada, 2016), there is an urgent need for research on linguistic markers of French-speaking teenagers with DLD.

2.1.4. The present study

We demonstrated in the previous sections that there is an urgent need for studies that identify reliable tasks that discriminate French-speaking adolescents with and without DLD, and even more for the Quebec population. In order to obtain a more comprehensive portrait of this population, research should combine typical tasks used in the two contexts where we find language assessments of teenagers with DLD, i.e., clinical and research settings. For instance, it has been shown, mostly in English, that metalinguistic skills associated with lexico-semantics and morphosyntax are impaired in DLD teenagers, but to our knowledge the discriminating ability of these tasks were never compared to tasks used in clinical settings, such as sentence repetition. The first objective of the present study is to examine the diagnostic accuracy of tasks used in research and clinical setting to discriminate between a group of French-speaking teenagers known to have DLD since childhood, and a group of typically developing pre-teens and teens. The tests and tasks selected for this study were taken from published and experimental materials with the goal to cover areas of language that have been identified as weaknesses in older children and adolescents with DLD. We examined which of these tasks provided the highest degree of accuracy in identifying adolescents with DLD. Adolescents with DLD tend to show heterogenous patterns of severity as a population and within individuals (Conti-Ramsden, 2008), with severity differences depending on the linguistic subdomains. Considering this and the lack of studies on French adolescents with DLD, we do not have clear expectations as to which tasks are likely to be the most discriminating or difficult for our population. Elin Thordardottir et al. (2011) showed that among ten language tasks, five accurately discriminated between French-speaking 5-year-olds with and without DLD: receptive vocabulary, receptive morphosyntax skills, nonword and sentence repetition and a following-directives task. Closer to the age of our group, the sentence repetition task has already been demonstrated to be discriminating in French-speaking children with DLD (Leclercq et al., 2014), as it is for elicited production of subject-verb number agreement the irregular verb for faire ‘to do’ (Franck et al., 2004) or past-tense production (Royle et al, 2018). A second objective of this study was to assess the linguistic subdomain that especially represent an area of weakness in French adolescents with DLD. To do so, we directly compared the tasks used in this study that yielded the best discriminating power for teens with DLD and investigated which one had the best diagnostic accuracy and to which linguistic subdomain this task corresponded. Despite morphosyntactic difficulties not being salient in spontaneous speech of young French-speaking children with DLD when compared to lexico-semantic skills (Elin Thordardottir & Namazi, 2007), targeted tasks could highlight morphosyntax as an area of weakness in French adolescents with DLD.

2.2. Methods

2.2.1. Participants

A total of thirty-seven French-speaking children and teenagers participated in this study, which is part of a larger research project on neurocognitive processing in DLD (see Courteau et al., 2019). The protocol was approved by the University of Montreal Research Ethics Board for educational and psychology research (CERES-15-070-D(4)). In accordance with the Declaration of Helsinki, all participants’ parents gave written consent for their child’s participation and participants themselves gave oral consent prior to the first experimental session. All had a hearing screening on the first day of assessment (500 Hz to 8000 Hz at 25 dB in at least one ear). Their mother tongue was French and was their language of instruction and daily use.

A group of seventeen teenagers with DLD (DLD group), including 10 girls, aged between 12 and 15 years (M = 14;01; SD = 0;72) served as the clinical target group, in relation to which the language measures' diagnostic accuracy was determined. The majority (n = 14) were recruited from a specialized private school for children and adolescents with learning disabilities in Montreal (Quebec, Canada) through a letter of invitation sent by the school speech-language pathologist (SLP) to the parents of students meeting the selection criteria. It should be noted that this school excludes children with disruptive behavior, possibly explaining why our group of participants with DLD includes more girls than boys. The other participants were recruited from a parent's association for children with DLD.

The inclusion criteria of DLD were a clinical diagnosis of DLD since childhood, functional impairments that meet the DLD definition as detailed by Bishop et al (2017), and persistent language impairments. All participants had a documented history of DLD and a complete SLP language evaluation (including discourse and pragmatic domains) resulting in a diagnosis. All teenagers of the DLD group had been diagnosed before kindergarten or during the first year of primary school, and maintained significant functional impairments needing adaptations to succeed in school. These were for the most part accommodations in regular classes or enrolment in special ones. Note that many participants had co-morbid disorders, such as ADHD and dyspraxia. These disorders do not preclude a DLD diagnosis (see Statement 9; Bishop et al., 2017). A study by Redmond and colleagues (2015) showed that ADHD co-morbidity with DLD–and TL development–does not increase children's errors on language assessment tasks such as sentence recall. However, the dominant clinical profile of these participants was the presence of persistent language difficulties as shown by their significantly lower scores as a group than the typical language group (TL) group on the Recalling sentences task (from the CELF-IVcnd-F, Secord et al., 2009, see Table 1). This task assesses semantic, morphological, and syntactic domains and has been shown to discriminate between typical and disordered teenage language development in English (Conti-Ramsden et al., 2001) and in French (Leclercq et al., 2014). Exclusionary criteria were the presence of associated biomedical conditions such as intellectual disabilities and autism spectrum disorder, and bilingual language acquisition.

Twenty participants with no history of language impairment (7 girls), aged between 7 and 14 years (M = 12.2; SD = 2.25) formed the typical language (TL) group. Their typical developmental status was established via a questionnaire filled out during an interview with their parents, and confirmed by our linguistic and cognitive tasks. None had any significant prenatal or perinatal complications, extended hospitalization, or serious illness.

Both groups were matched on non-verbal abilities using 4 tasks within the Cognitive Experiments IV v2 pack of the Presentation® software (Version 18.0, Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com). Visuospatial working memory was assessed with the forward and backward Corsi Blocks tasks (Corsi, 1972) and by a delayed match-to-sample task of non-verbal stimuli (Daniel et al., 2016) with delays of 1 or 5 seconds. Participant characteristics for both groups are presented in Table 1. To compare groups statistically, we used Brunner-Munzel tests (Brunner & Munzel, 2000) as recommended by Rietveld and van Hout (2015) for skewed data with small sample sizes. The Brunner-Munzel is a robust nonparametric test that checks the stochastic superiority of a group, expressed by the Brunner-Munzel statistic (tbm), a p-value and a Common Language Effect Size (CLES), indicating in our case the probability of a random observation from the TL group being larger than a random observation from the DLD group, with 0.5 being at chance. Differences between groups were found in age (DLD > TL), schooling (DLD > TL), and on the Recalling Sentences test (DLD < TL). Note that participants were meant to be matched on the age variable, but that recruitment was halted due to the COVID pandemic. Since DLD group is significantly older than the TL group, higher TL group scores cannot be attributed to age. See Figure 1 for a display of the age variable distribution.

Table 1

Participant characteristics with comparisons between groups made by the Brunner-Munzel test.

	TL group (N = 19)		DLD group (N = 17)		Brunner-Munzel tests		
	Mean	SD	Mean	SD	t_{bm}	p-value	CLES
Age	12.21	2.25	14.09	0.77	3.39	0.002	0.24
School	5.9	2.20	7.53	0.51	2.30	0.03	0.30
Recal	67.75	9.72	54.76	7.57	5.55	< .0001	0.85
Corsi–F	5.55	1.76	5.56	1.55	0.12	0.91	0.51
Corsi–B	5.60	1.76	4.94	1.06	1.18	0.25	0.61
DMTS–1s	0.89	0.11	0.88	0.10	0.48	0.63	0.55
DMTS–5s	0.82	0.15	0.84	0.13	0.24	0.81	0.47

Note. Chronological age (Age) and schooling (School) are expressed in years. Recalling Sentences (Recal) CELF-IVcnd-F scores are untransformed, Corsi Blocks scores reflect forward (Corsi–F) and backward (Corsi–B) untransformed spans, and delayed match-to-sample represent percentage accuracy for 1 second (DMTS–1s) and 5 second (DMTS–5s) delays.

2.2.2. Procedure

Experimental sessions took place in a quiet room either at the participants' high school or at the University of Montreal in the fourth author's lab. Participants were individually tested for two 2–2.5 hour sessions where, in the first hour, they participated in an ERP experiment (Courteau et al, in preparation). Testing was conducted by a Quebec-accredited SLP, i.e., the first author, or trained research assistants. All experimenters had French as their native language. The tasks used in this study can be classified into two categories and will be briefly described in the following order: 1) those commonly used by SLPs in clinical settings to assess the language skills of Quebec French adolescents, and 2) those used in DLD research. See supplementary materials for a detailed presentation of tasks.

Three tasks were selected from the CELF-IV$^{cnd\text{-}F}$ French version standardized among Quebec-French speakers (Secord et al., 2009) and were administered as recommended by the manual. We used the Recalling Sentences task, where participants repeated orally-presented sentences, the Word Classes task, which assessed the ability to understand lexico-semantic relationships between orally presented words by choosing the two words that go together (receptive subtask) and explaining this relationship (expressive subtask). Third, the number repetition tasks consisting of a forward and backward digit span were used to evaluate phonological working memory. This memory was also evaluated using the non-standardized French Quebec nonword repetition Courcy task (Elin Thordardottir et al., 2011) which consists of 40 words. Total repeated phonemes, with a maximum of 180, was used as participants' score. We chose the EVIP (a Canadian-French adaptation of the PPVT, Dunn et al., 1993) to evaluate receptive vocabulary. Expressive vocabulary was assessed with an action (verb) naming task from the French version of the fLEX test (see Pourquié et al., in revision, for details). We evaluated subject-verb number agreement skills with two tasks from fLEX which included 20 verbs that had an audible agreement number cue (e.g., il rugit [ilʁyʒi], 'he roars' vs. ils rugissent [ilʁyʒɪs], 'they roar'). The expressive task assessed subject-verb number agreement on verbs via elicited sentence production, and the receptive task assessed their understanding. The fLEX test is a research tool and does not provide norms, but we selected it for its thorough control of linguistic cues.

As tasks commonly used in research on language-acquisition, we first used two grammaticality judgment tasks based on an alien-learning paradigm (Courteau et al., 2013) where an alien practices French and sometimes makes mistakes. Participants listened to pre-recorded sentences while looking at pictures and judged if sentences were correct or not. Participants' grammaticality judgments were quantified with A-scores, a biais-adjusted measure of sensitivity which includes the participant's ability to correctly classify presented sentences as containing an error or not (A'-score, corrected version, Zhang & Mueller, 2005). A-scores of 1 reflect perfect discrimination, and 0.5 chance levels. In the first task, adapted from Poulin et al.

(2015), participants looked at pictures while listening to 16 sentences that were either correct (n = 4) or contained errors targeting the noun phrase (n = 12). Errors included auditory-visual lexico-semantic mismatches on nouns (n = 4) and morphosyntactic gender-agreement errors on determiners or adjectives (4 each). The second task was run during an ERP experiment (Courteau et al., in preparation). Participants looked at pictures while they listened to 300 grammatical sentences that were either a match or contained errors that targeted the verb (150 each), and judged if the visuo-auditory pairs were a match or not by using a button press. Lexico-semantic errors were created using a verb that did not match the depicted action (n = 30). Subject-verb number-agreement errors were created by varying the number of visually presented agents and morphosyntactic number cues in the auditory stimuli. This was done using either subregular and irregular verbs whose number cue was audible on the verb ending (n = 60; e.g., visual: [A LION ROARS], auditory: En soirée, ils *rugissent [ilʁyʒɪs] dans la savane ‘In the evening, *they roar in the jungle’) or with regular agreement morphophonology, where the plural number cue ‘s’ [z] was realized through liaison between the pronoun plural form and verb’s onset which was always a vowel (n = 60). We also assessed visuospatial working memory trough the Corsi Blocks tests (Corsi, 1972) and delayed match-to-sample tasks of non-verbal stimuli (Daniel et al., 2016) as described in the participant section. Interrater reliability was calculated on all tasks that involved a verbal response for every participant with DLD and four of the TL participants. Based on Krippendorff (2004), interrater reliability percentage of agreement was at a minimum 95% and a maximum of 100% across subtasks (Duquette et al., 2020).

In total, we administered 20 subtasks which generated 24 scores per participant. Three score types were produced depending on the task: 13 raw scores corresponding to the untransformed total number of subtask items successfully completed by the participant, five percentile rank scores derived from the same-age norm group for standardized subtasks, and six A-scores for the grammaticality judgment subtasks. Note that one participant didn't complete the Word Classes and Corsi bloc tasks (DLD-01, 20/24 scores), and another didn't complete the DMTS tasks (TL-06, 22/24 scores). We listed in Table 2 to which linguistic or working memory subdomains each subtask corresponded according to the tests' manual, when available, or based on the literature as presented in the introduction section (for a detailed list of score types and subdomains associated with each subtask see supplementary material Table1).

Table 2

Linguistic and working memory subdomains assessed in this study with their associated subtasks

Subdomains	Subtasks
Lexico-semantics	Word Classes, EVIP, fLEX action (verb) naming
Metalinguistics: semantics	Grammaticality judgments on lexical-semantics for nouns and verbs
Morphosyntax	fLEX inflected verb production and comprehension
Metalinguistics: morphosyntax	Grammaticality judgments for gender agreement in noun phrases, and subject-verb number agreement in sentences containing regular or irregular verbs.
Phonological working memory	Forward and backward digit span, Nonword repetition
Visuospatial working memory	Forward and backward Corsi blocks, DTMS–1s and DMTS–5s
Lexico-semantics, morphosyntax and phonological working memory	Recalling Sentences

2.2.3. Analyses

2.2.3.1. Variable selection

We applied a variable selection procedure to identify which ones of the 24 subtask scores had the potential to discriminate between groups and to avoid multicollinearity problems, a common concern with multivariable models. Using RStudio version 1.4.1103-4 (RStudio Team, 2020), we calculated the information gain (IG; Azhagusundari & Thanamani, 2013) for each variable, and rejected those that had a null IG. Next, based on test specifications and IG, we eliminated variables that reflected pairs of scores that originated from the same subtask, such as raw and percentile equivalents of the same measurement, retaining the score with the best IG.

2.2.3.2. Correlational analyses

We performed a correlational and cluster analyses to explore statistical relationships between scores and to verify that the correlated score clusters corresponded to linguistic subdomains that the underlying subtasks were intended to measure. The variables were Group, Age, and the previously-selected scores. Considering the small sample size and scores with varied distributions, Spearman rank correlations were used to provide robust estimates of the variable associations. The interpretation thresholds for Spearman coefficients were those suggested in Akoglu (2018) for psychology research: +/- 0.1-0.3, weak; +/- 0.4-0.6, moderate; +/- 0.7-0.9, strong; +/- 1, perfect. We also produced a clustered dendrogram to further analyze the hierarchical relationships between scores that displayed similarities. The hierarchical clustering was determined using Ward’s method (Murtagh and Legendre, 2014).

2.2.3.3. Group comparisons

Multiple comparisons using Brunner-Munzel tests (Brunner & Munzel, 2000) implemented in the *bunnermunzel* R package (Hui et al., 2020) were applied to assess the difference between the TL and DLD groups. In our case, this test estimates the probability that a participant randomly drawn from the TL group will have a higher score than a participant randomly drawn from the DLD group. We applied a Bonferroni-Holm adjustment to the resulting p-values to control the false discovery rates (Abdi, 2010). We report the confidence intervals (CI) associated with the resulting p-values.

2.2.3.4. Optimal cut-off scores

The subtasks were considered as tests with threshold scores, i.e., cut-off scores, and two possible outcomes: below the cut-off scores, the participant is assumed to have DLD and above, to have TL development. To analyze the discriminatory ability of the subtask's scores, we identified optimal cut-off scores based on our sample data and calculated measures of diagnostic accuracy. The optimal cut scores were estimated by a bootstrap procedure that randomly resampled our groups but with replacements, a thousand times, using the *multi_cutpointr* function of the R package *cutpointr* (Thiele & Hirschfeld, 2020). The selected cut point for each variable was the point that maximized the sum of sensitivity and specificity. To mitigate sample bias, the cut point was recalculated as the midpoint between the optimal cut point and the next lowest score. Ties were resolved by returning the mean of conflicting cut points. Note that the recommended cut-off scores for subtasks from standardized tests, generally start at -1SD, or about the 16th percentile (Conti-Ramsden et al., 2001).

We report several measures of diagnostic accuracy (Hajian-Tilaki, 2013) associated with the selected cut points. Sensitivity is the true positive rate, which for our study means the proportion of participants with a documented DLD who are identified as such by the subtask. Specificity is the true negative rate, the proportion of participants with TL development who are identified as such by the subtask. For these two measures, a proportion above 90% is considered a good discriminant, between 80% and 89% fair, and below 80% unacceptably low (Plante & Vance, 1994). These two measures can be combined in an index: the likelihood ratio. A positive likelihood is the ratio of true positives to false positives; higher values indicate more informative tests. A ratio of 10 is considered strong (Jaeschke et al., 1994) and indicates that the likelihood of having a DLD would be 10 times higher if the participant's score is below threshold than if it was above. Inversely, a negative likelihood is the ratio of false negatives to true negatives; values close to 0 indicate more informative tests: a ratio of 0.1 is considered strong. Other qualitative terms used to describe positive/negative likelihood ratios are the following: 5-10/0.1-0.2, moderate; 2-5/0.2-0.5 small and sometimes important; 1-2/0.5-1 small but rarely important (*ibid*). The receiver operating characteristic (ROC) curve is a plot of the true positive rate (sensitivity) to false positive rate (1 - specificity) for all possible cut-off scores; the derived area under the ROC curve measurement (AUC) can be interpreted as such: 0.5 is no better than chance, 0.5-0.7 equals poor discrimination, 0.7-0.8 is acceptable, 0.8-0.9 is excellent and over 0.9 is considered outstanding classification. For all measures associated with the optimal cut-off scores, we calculated their 95% confidence interval (CI), representing percentiles 2.5 and 97.5, based on the distribution produced by 1000 bootstrap iterations.

2.2.3.5. Multivariable analysis

To provide a broader picture of the linguistic subdomains in which DLD teenagers exhibit weaknesses while accounting for correlations between scores, we fitted a multivariable logistic regression model to predict the group. The predictor variables were the selection of subtasks previously identified as discriminating between TL and DLD, minus the Recalling Sentences score, which we removed from the set of predictors as the underlying subtasks span several linguistic subdomains. Collinearity of the remaining six predictor variables was first assessed using the *findLinearCombos* function from R's *caret* package (Kuhn, 2009), which did not reveal multicollinearity problems. The same method applied to centered and normalized scores also did not reveal multicollinearity problems. Since a regular logistic regression procedure resulted in near-perfect separation that prevented us from producing the relevant statistics, we used a regularized logistic regression procedure R's *glmnet* package (Hastie et al., 2016). This method required that all the variables were on the same scale prior to model fitting. All subtasks with non percentile scores were centered and scaled, which refers to subtracting the mean and dividing by the standard deviation. Subtasks with percentile scores were converted to *z* scores using the normal distribution. The lambda-regularization parameter, used to determine how strict the regularization is, was set by a leave-one-out cross-validation procedure. We selected the smallest lambda value that minimized cross-validated classification errors. The relative contribution of the variables to the model was estimated with the permutation method (Altmann et al., 2010) implemented in the *vip* R package (Greenwell & Boehmke, 2020). This permutation method measures the difference to a performance metric when the values of a predictor variable are shuffled, in our case the AUC when a subtask's scores are shuffled across all participants thus making this variable uninformative. We report the resulting coefficients across 100 repetitions of the permutation procedure, to rule out accidental patterns in the shuffled data, along with the relative importance of the variables.

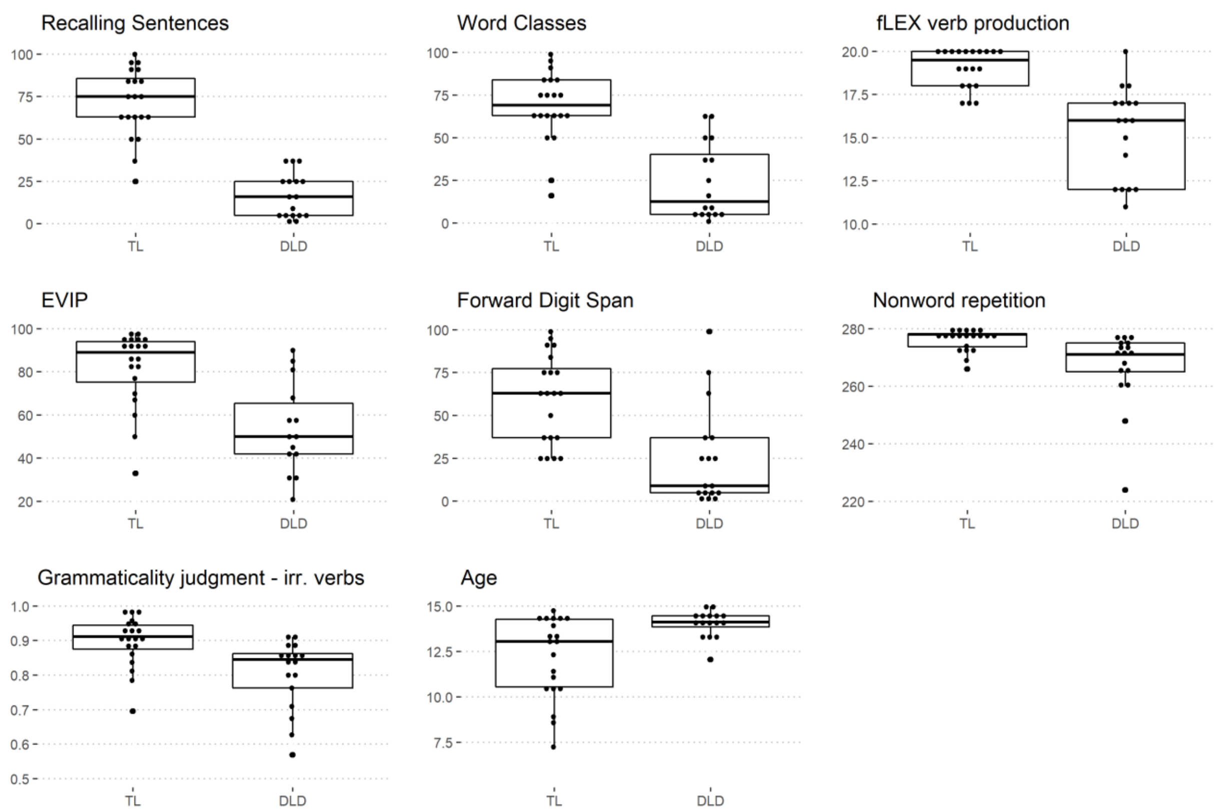
2.3. Results

2.3.1. Variable selection

Out of 24 possible scores, 17 had a null IG (IGs for all subtask scores are presented in supplementary materials, Table 2). Of the remaining nine scores, two pairs of scores stemmed from identical or similar subtasks. Both the Recalling Sentences task's percentile score (IG = 0.52) and raw score (IG = 0.27) had a positive IG, so the score with the lower IG was removed from the selection. The Word Classes production subtask score was removed given that it is dependent on the Word Classes comprehension score, which was kept in the selection because of its higher IG. The Digit Span total score was removed because it is the scaled sum of the forward and backward scores from the same task. The final variable set was composed of 4 age-based percentile scores (Recalling Sentences, Word Classes, Forward Digit Span and EVIP), 2 raw scores (fLEX irregular verb production: maximum 20 verbs, and Nonword repetition: maximum 280 phonemes) and one *A-score* (Grammaticality judgment on irregular verb agreement). Figure 1 shows the distribution of the final selection of scores.

Figure 1

Results for both DLD and TL groups on 7 discriminating subtasks and age.



Note. The y axis indicates the score. Scores are presented as percentile scores except fLEX verb production (max: 20 target verbs), Nonword repetition (max: 280 target phonemes), Grammaticality judgment of irregular verbs (A-score, 1 being perfect, 0.5 indicating chance). The Age scale is expressed in years.

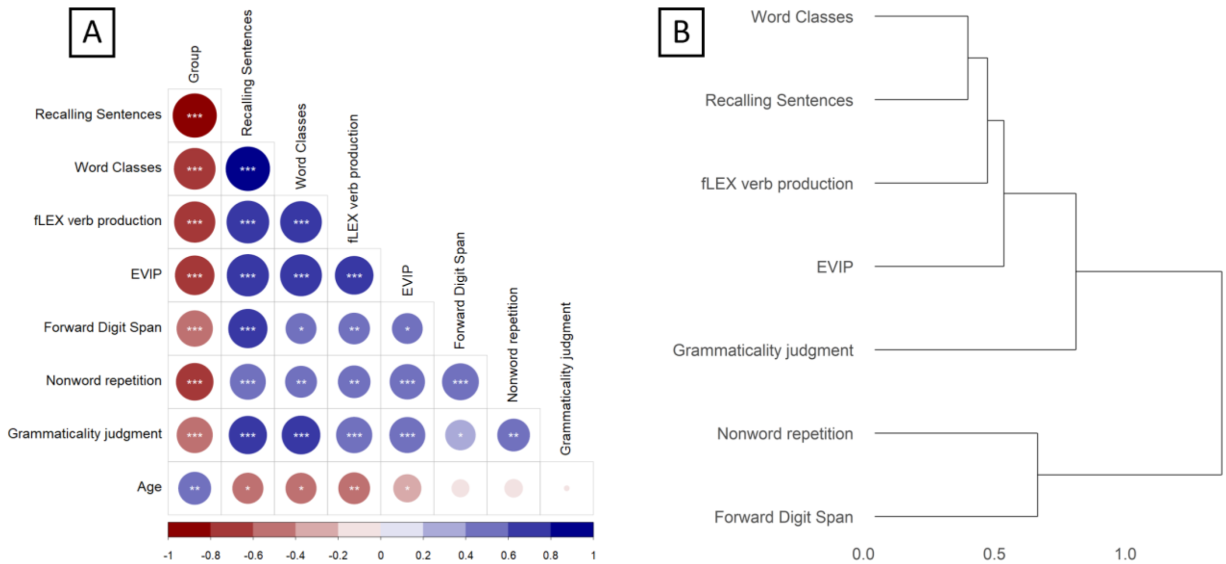
2.3.2. Correlational analyses

Figure 2A shows the correlation matrix plot for Age, Group, and the seven selected subtask scores. All selected test scores had a statistically significant correlation with Group at a *p*

< 0.001 level, with Spearman coefficients ranging from *rs* = -0.56 (Grammaticality judgment on irregular verbs) to *rs* = -0.83 (Recalling Sentences), indicating moderate to strong associations. Correlations between test scores were moderate to strong (*rs* = 0.39 to 0.85) and all reached statistical significance at a $p < 0.05$ level. Forward Digit Span (*rs* = 0.39 to 0.66) and Nonword repetition (*rs* = 0.39 to 0.58) displayed relatively weaker associations with the other test scores. The dendrogram in Figure 2B shows the variable's hierarchical structure as a way to assess consistency between variable groupings (clusters) and linguistic and working memory subdomains. The test scores formed a large cluster, separate from the Age variable, and divided between two subclusters: phonological working memory subtasks (Forward Digit span and Nonword repetition) and language subtasks. The language subcluster was itself divided between the linguistic subtasks (Word Classes, Recalling Sentences, fLEX verb production and EVIP) and the metalinguistic subtask (Grammaticality judgment on irregular verbs). See supplementary materials Figure 1 for a correlation plot with exact Spearman coefficients.

Figure 2

Correlation plot for selected subtasks, with group and age (A). Hierarchical clustering for subtask scores displayed in a dendrogram (B)



Note. Subtask Grammaticality judgment of irregular verbs. A. Coefficients and levels of statistical significance are based on Spearman rank correlations. B. The x axis is the value of the Ward minimum variance criterion, used as a distance measurement.*p < 0.05, **p < 0.1, ***p < 0.001.

2.3.3. Groups comparisons

We examined group differences on the seven discriminating tasks as seen in Table 3. The TL group showed significantly better performance in all subtasks as seen by their higher group means and as demonstrated by *p* values below 0.001. Three subtasks were found to be the most discriminating with a CLES below 0.1: Recalling Sentences, Word Classes and fLEX production of irregular verbs. The other tasks had CLES between 0.11 and 0.18.

Table 3

Group comparisons on seven discriminating subtasks

Subtasks	Group mean (SD)		Brunner-Munzel Tests		
	TL	DLD	t_{bm}	*p*	CLES [CI]
Recalling Sentences	71.3 (20.3)	16.5 (13)	-28.88	<0.001	0.02 [-0.02, 0.05]
Word Classes	67.8 (21.2)	24 (22.4)	-10.8	<0.001	0.07 [-0.01, 0.15]
fLEX verb production	19 (1.1)	15.3 (2.7)	-8.27	<0.001	0.09 [-0.01, 0.19]
EVIP	81.8 (17.5)	46.3 (25.1)	-7.27	<0.001	0.11 [0.01, 0.22]
Forward Digit Span	59.9 (25.7)	25.2 (29)	-4.34	<0.001	0.18 [0.02, 0.33]
Nonword repetition	276.2 (3.8)	266.7 (13.4)	-5.61	<0.001	0.15 [0.03, 0.28]
Grammaticality judgment	0.9 (0.1)	0.8 (0.1)	-4.64	<0.001	0.18 [0.04, 0.32]

Note. Subtasks Grammaticality judgment of irregular verbs. Brunner-Munzel p-values are presented with Bonferoni-Holm adjustments and common-language effect sizes (CLES) with their confidence intervals (CI).

2.3.4. Optimal cut-off scores

We identified optimal cut-off scores and related measures of diagnostic accuracy, as seen in Table 4. Recall that four of our seven more discriminating subtask scores are age-based percentile ranks. The recommended cut-off score for standardized tests is typically -1SD, or about the 16th percentile. Results showed that the optimal cut-off score for the Forward Digit Span task was at the 17th percentile, with the value 16 being part of the CI, and thus similar to what is recommended by the test. We found much higher cut-off scores for Recalling Sentences (43.5th), Word Classes (62.5 th) and the EVIP (59 th). For these subtasks, the CIs did not include the 16th percentile, indicating that a cut-off score of 16 is unlikely based on our sample.

Table 4

Bootstrap estimated optimal cut-off scores and their derived measures of diagnostic accuracy.

Subtask	Optimal Cut-off	Sensitivity	Specificity	PosLH	NegLH	AUC
Recall.	43.5	1	0.9	10	0[a]	0.98
	[31, 50]	[0.92, 1]	[0.76, 1]	[4.25, ∞]	[0, 0.08]	[0.94, 1]
Word C.	62.5	0.94	0.8	4.69	0.08	0.93
	[20.5, 62.75]	[0.68, 1]	[0.64, 1]	[2.57, ∞]	[0, 0.34]	[0.82, 0.99]
fLEX	17.5	0.82	0.85	5.49	0.21	0.91
	[16.5, 18.5]	[0.57, 1]	[0.64, 1]	[2.54, ∞]	[0, 0.43]	[0.8, 0.99]
EVIP	59	0.76	0.9	7.65	0.26	0.89
	[47.5, 90.5]	[0.61, 1]	[0.55, 1]	[2.14, ∞]	[0, 0.42]	[0.77, 0.97]
F. Digit	17	0.53	1	∞[a]	0.47	0.82
	[15, 50]	[0.37, 1]	[0.61, 1]	[2.21, ∞]	[0, 0.63]	[0.66, 0.95]
Nonword	277.5	1	0.55	2.22	0[a]	0.85
	[268.5, 277.5]	[0.53, 1]	[0.42, 1]	[1.73, ∞]	[0, 0.5]	[0.71, 0.95]
Gram. J.	0.88	0.82	0.75	3.29	0.24	0.82
	[0.86, 0.92]	[0.68, 1]	[0.44, 0.95]	[1.75, 15.84]	[0, 0.4]	[0.67, 0.94]

Note. PosLH.: Positive likelihood; NegLH.: Negative likelihood; Recall.: Recalling Sentences; ∞.: Infinite; Word C.: Word Classes; fLEX: fLEX verb production., F. Digit: Forward Digit Span; Nonword.: Nonword repetition; Gram. J.: Grammaticality Judgment on irregular verbs. Optimal cut-off scores estimated from 37 participants (n = 36 for Word Classes). Derived measures of diagnostic accuracy are listed with 95% confidence intervals (CI) representing percentiles 2.5 and 97.5 with 1000 bootstrap iterations.[a] For NegLH, 0 means perfect identification of TL participants and for posLH, ∞ means perfect identification of DLD participants.

Related to our optimal cut-off scores, Recalling Sentences showed the highest sensitivity and specificity, above 0.90, followed by Word Classes with good to fair sensitivity and specificity, respectively 0.94 and 0.8, and fLEX with fair levels on both measures, 0.82 and 0.85. All other tasks exhibited measures under 0.80 on either one of these measures, indicating unacceptably low accuracy. Regarding likelihood ratios, only Recalling Sentences, Word Classes and fLEX subtasks revealed strong, moderate or near moderate effects on both positive and negative ratios. It is not surprising that these subtasks also revealed the highest AUCs, of over 0.9, or 0.8 when considering their CIs, reaching outstanding or at least excellent classification of participants. The other subtasks had AUCs between 0.82 and 0.89, but because their AUCs dropped under 0.8 in the CIs, their accuracy only reached an acceptable classification. ROC plots for all seven subtasks are available in supplementary materials, Figure 2).

2.3.5. Multivariable analysis

We fitted a regularized logistic regression model with group classification (TL or DLD) as a dependent variable, and six subtask scores (our selection of seven subtask scores, minus Recalling Sentences, as it taps lexico-semantics as well as morphosyntactic linguistic domains) as independent variables. The resulting coefficients are shown in Table 5 along with relative variable importance. Since the variables were set to the same scale before model fit, a coefficient further from zero indicates a stronger contribution of the variable. The coefficients for Nonword repetition and Grammaticality judgment shrunk to zero, indicating that these variables were eliminated by the regularization procedure. Of the remaining 4 subtask scores, fLEX verb production showed the largest contribution; including this score improved the AUC of the model by an average of 0.2, as opposed to including a randomly permuted vector containing the same values. The model improvement was smaller for Word Classes (mean AUC gain = 0.028) and EVIP (0.029) and almost null for Forward Digit Span (0.001). When used to classify the participants between TL and DLD groups, the final model had a sensitivity of 0.88, a specificity of 1 and an AUC of 0.98. Of 36 participants with no missing values, the model accurately classified 34 and produced 2 false negatives (two participants with DLD classified as being the TL group).

Table 5

Coefficients and variable importance for the regularized logistic regression model

Subtasks	Coefficient	Var. contribution (SD)
fLEX verb production	-1.00	0.200 (0.067)
EVIP	-0.424	0.029 (0.018)
Word Classes	-0.423	0.028 (0.020)
Forward Digit Span	-0.169	0.001 (0.010)
Nonword repetition	0	0 (0)
Grammaticality judgment irregular verbs	0	0 (0)

Note. Coefficients and variable importance produced with lambda parameter previously set by a leave-one-out cross-validation procedure. Variable contribution indicates the mean difference in the model's area under the curve (AUC) when the variable is permuted; results shown are for 100 permutations.

2.4. Discussion

The present study first aimed to identify reliable tasks used in research and clinical settings that discriminate French-speaking adolescents with DLD from their TL peers. Based on 20 subtasks administered to 37 older children and teenagers with and without DLD, we found seven tasks from different language domains that displayed high levels of diagnostic accuracy. A second objective was to assess which linguistic subdomain(s) more specifically represented areas of weakness in our DLD participants. To do so, we compared our most discriminating subtasks directly to see which one(s) contributed the most to identify French teenagers with DLD. We found that the irregular verb production subtask assessing morphosyntactic skills contributed more to the model's diagnostic accuracy than subtasks assessing lexico-semantics and phonological working memory, revealing morphosyntax as a special area of weakness in French-speaking teenagers with DLD.

2.4.1. Discriminating tasks for French-speaking adolescents with DLD

Of the 20 subtasks administered to participants, seven subtasks were found to be informative about the group to which the participants belonged. Using clusters to examine the relation between these subtasks (Figure 2), we found that a first cluster included phonological working memory subtasks and a second one language subtasks. Interestingly, the language cluster was divided in two subclusters: one targeted metalinguistic skills and the other one language production or comprehension subtasks. We found moderate correlations between subtasks of the first (grammaticality judgment on irregular verb agreement) and second subcluster (Recalling Sentences, Word Classes, fLEX irregular verb production and the EVIP). This result suggests that, as suggested by Van Kleeck (1982), metalinguistic tasks call for skills that are complementary to those assessed by language comprehension and production tasks in older children and teenagers. Scores on the four language subtasks were strongly correlated, except between receptive vocabulary (EVIP) and fLEX irregular verb production subtasks, which showed a moderate correlation. One interpretation of this finding is that the EVIP is assessing the comprehension dimension whereas fLEX is assessing the production one. However, two studies on the dimensionality of language have not found evidence to support comprehension and production as two different dimensions (Lonigan & Milburn, 2017; Tomblin & Zhang, 2006). Furthermore, we found that other tasks assessing these two dimensions were strongly correlated. A more compelling explanation is that these two subtasks are probably the two most representative ones of their respective domains and therefore tap into different linguistic skills. This is coherent with Tomblin and Zhang (2006), who showed that lexico-semantics and (morpho-)syntactic subdomains are to be considered distinct in older children.

The TL group performed better than the DLD group on all seven subtasks. The superiority of the TL group, younger on average than the DLD group, could partially be accounted for by the fact that their percentile scores were age-standardized for the EVIP, Recalling Sentences, Word Classes and Forward Digit Span subtasks. However, the TL scores on the experimental tasks (fLEX irregular verb production, nonword repetition and grammaticality judgment) were also higher, even though they were not age-standardized. In the case of verb production and the metalinguistic tasks, both targeting irregular verb agreement, this suggests that this type of agreement is well developed in (pre-)teens with TL but clearly impaired in DLD participants. This finding is in line with Rose and Royle (1999) and Franck et al. (2004). As these latter authors did, we found 5% of errors on average in the TL group (i.e., 1 error on 20 verbs) on irregular verb production. However, looking at individual score distributions (Figure 1), half of the TL participants actually made more than just one error. Our results suggest that typically developing French-speaking children and adolescents have not yet fully acquired the production of irregular verb agreement. Our results thus suggest that small numbers of errors in irregular verb agreement elicitation tasks should not be considered indicators of DLD in children 8 to 14 years old.

We calculated the optimal cut-off scores that best classified our participants in our two groups for each of the seven discriminating tasks. Recommended cut-off scores for standardized subtasks typically start at -1SD (or the 16th percentile) for mild language impairment. We found one score close to the recommended cut-off score for the CELF Forward Digit Span task (17th percentile). However, we found much higher cut-off scores for other standardized subtasks assessing linguistic skills, including the EVIP (59th percentile), Recalling Sentences (43.5th percentile), and Word Classes (62.5th percentile), with the 16th percentile also missing from the bootstrapped CIs. In a nutshell, all our participants exceeded the expected performance, as our optimal cut-off scores were close to the average scores in published norms, i.e., 50th percentile. What could explain these surprising results? A first interpretation would be that our sample was composed of particularly high performing participants for their age. This is implausible because most of our adolescents with DLD had important language impairments as evidenced by their attendance at a school with special accommodations. A second explanation would be that the published norms were conducted on surprisingly low-performing groups, which is unlikely, at least for the CELF, given that it was based on a considerable number of children; 520 francophones from Quebec aged 4 to 16. Another explanation for our high cut-off scores could be that the French CELF test, which 3 of our 4 standardized subtasks were taken from, has poor psychometric properties resulting in inadequate recommended cut-off scores; however, the English version of this test was identified as one of the recommended tests to evaluate language, based on its good psychometric quality (Denman et al., 2017). The most compelling explanation for these results would be the problematic adaptation and translation of language stimuli in these tests, which were not focused on specific linguistic constructs important for assessing Quebec French language development. Along with Godard and Labelle (1995) and Elin Thordardottir et al. (2011), our finding is another demonstration of how French versions of English tests used by Quebec SLPs underestimate language difficulties in teenagers with DLD. These results also support the argument that low scores on standardized tests should not be a criterion to assess disorder severity (Bishop et al., 2017; Breault et al., 2019; Tessier & Valade, 2017).

Based on our optimal cut-off scores, we calculated diagnostic accuracy statistics. Of the seven subtasks we selected, three were found to have outstanding discriminating ability, with AUCs above 0.90, and maintaining excellent AUC values in their CI. These subtasks also had good to fair sensitivity and specificity, as well as strong, moderate or near moderate effects on both positive and negative likelihood ratios. Unsurprisingly, Recalling Sentences discriminated best between our teenage participants with DLD versus TL, with a AUC of 0.98. This task had been proven to be a powerful diagnostic tool with French children aged 5 (Elin Thordardottir et al, 2011) and 7 to 12 (Leclercq et al, 2014), and we now can confirm that it is still highly relevant when assessing DLD in 14-year-olds. Our second-best diagnostic task was Word Classes (AUC = 0.93). This result is consistent with studies by McGregor et al. (2013) and Rice and Hoffman (2015), which also found lexico-semantic deficits in teens with DLD, and suggests that future research should consider including lexical-semantic skills as a marker for DLD. Furthermore, this result highlights that when assessing lexico-semantic skills in teenagers, the use of multi-word tasks such as Word Classes is more accurate in diagnosing DLD compared to a single-word task such as the EVIP. Indeed, even if the EVIP's AUC (0.89, see Table 4) was close to that of Word Classes, the EVIP's AUC was below 0.80 when looking at the CI, which is considered only acceptable, whereas the Word Classes' CI AUC remained excellent, and over 0.80. The third subtask with the best AUC was fLEX irregular verb production. This finding suggests that irregular verb morphology production deficits are a salient characteristic of French teenagers with DLD. Elin Thordardottir et al (2011) found that receptive morphosyntactic assessment, but not production, was one of the five best tasks that accurately discriminated French-speaking 5-year-olds with and without DLD. This difference might be because, as they age, children with DLD will better master comprehension, but maintain difficulties in verb production.

The two tasks assessing phonological working memory, Forward Digit Span and nonword repetition, had lower diagnostic accuracy. Even in their AUCs were above 0.80, which is considered excellent, their CIs were only acceptable. Two studies that assessed phonological working memory with either number or nonword repetition have found lower performance in teens with DLD when compared to their TL peers, which is in line with our results (Arslan et al., 2020; Ebbels et al., 2012). However, they didn't analyze diagnostic accuracy as we did. Of typical tasks used in research on DLD, only one discriminated between our groups, namely grammaticality judgments on irregular verbs. This finding shows that this kind of morphosyntactic metalinguistic task taps into deficits in teenagers with DLD and is consistent with many studies that found similar results (Rose & Royle, 1999; Maillart & Schelstraete, 2005; Miller et al., 2008; Noonan et al., 2014; Poulin et al., 2015; Haebig et al., 2017). However, these studies didn't evaluate tasks diagnostic accuracy. Our results shows that this morphosyntactic metalinguistic task's diagnostic accuracy was low, as illustrated by the AUC that was only acceptable when looking at the CI, but recall that our task was presented in the context of an ERP study (Courteau et al, in preparation) and not in a typical clinical setting. Given the extensive evidence of impaired morphosyntactic metalinguistic skills in DLD, future studies should focus on assessing the diagnostic accuracy of these types of tasks in clinical settings.

2.4.2. Linguistic subdomains as areas of weakness

Using a multivariable model, we were able to explore the relative contribution of each subtask to the model's total AUC, which was 0.98 and corresponded to an outstanding classification. Only three subtasks were considered to contribute significantly to it. The subtasks assessing lexico-semantics–Word Classes and EVIP–contributed respectively 0.028 and 0.029 AUC to the model, whereas fLEX irregular verb production contributed 0.20 AUC. These results strongly imply that impairment in morphosyntax is a more discriminant characteristic of French-speaking adolescents with DLD as opposed to impairments in lexical semantics. This result is consistent with Leclercq et al. (2014) who compared participants' subscores for morphosyntactic and lexical-semantic errors in a sentence recall task in French-speaking children aged 7 to 12. They found, based on a principal component analysis, that morphosyntactic subscores provided the largest loadings on the first factor, while lexico-semantic subscores were associated with the second factor. Morphosyntactic impairments being a salient characteristic of DLD is also in line with many theoretical accounts of this disorder (for a review see Pourquié et al., in revision), including the Procedural Deficit hypothesis (Ullman & Pierpont, 2005). Elin Thordardottir and Namazi (2007) did not find morphosyntactic difficulties to be salient in spontaneous speech of 5-year-old children with DLD, when compared to lexico-semantic skills. In our view, their finding is due to the fact that 1) morphosyntactic skills are under-informative of language development in preschool children as they are still in development, at least when focusing on irregular verb agreement and 2) spontaneous utterances did not provide contexts in which they could target deficits specific to DLD (see also Royle et al, 2018). Our results are also coherent with Leonard (2014) who identified morphosyntax as an area of weakness in DLD across languages. Together with Elin Thordardottir et al. (2011), our findings offer clear evidence that children with DLD's linguistic skills change and evolve with age. Indeed, these authors identified two of our same subtasks, EVIP and nonword repetition, as the combination that provided the best diagnostic accuracy for 5-year-olds with an AUC of 0.98, whereas these tasks had little or no contribution to our adolescent data model. Our results carry implications for publishers of French tests targeting adolescents, who should consider including a subject-verb number agreement production task focusing on sub-regular and irregular verbs.

Based on irregular verbs agreement skills, the present study demonstrated that morphosyntactic impairments are a reliable marker of French DLD in teenagers, similar to what has been found in English for younger and older children and in school-aged French children. In order to investigate the scope of morphosyntax as a marker of DLD in French, future studies should compare several morphosyntactic skills beyond number agreement, and ensure that the psychometric properties of their tests are valid.

2.4.3. Study limitations

There are three main potential limitations to this study. A first limitation concerns our sample size. Considering our small sample, outliers could have caused accidental characteristics to be confused with actual trends. We used a robust statistical methodology to mitigate this problem. Participants in the control group were younger on average than those in the DLD group, and also covered a broader age range. Despite this, we found seven subtasks where their performance was superior to that of teenagers with DLD. However, if we had had a control group of the same age as those in the DLD group, perhaps additional tasks in our selection would have discriminated groups more robustly. A third potential issue concerns the selection of tests and tasks used in this study, which were not always adapted to the older age of our participants. Indeed, both groups performed almost at ceiling on many experimental tasks, partly explaining why only seven of 20 subtasks were found to be relevant. This was the case for the nonword repetition task, which was designed for kindergarteners. Even if it discriminated between groups, we recognize that a nonword repetition tasks best suited for an older population could have reached better diagnostic accuracy; for example a task with more complex syllable structures. A fourth limitation of this study is that we removed the Recalling Sentence task from the multivariable analyses. Following Leclercq et al., (2014), we could have generated sub-scores reflecting different skills and included them in the model.

2.5. Conclusion

Our study contrasted two types of language assessment tasks, namely clinical and research tools, that, to our knowledge, have not yet been directly compared in teenagers. This research can be considered as a first step towards identifying psycholinguistic markers of French-speaking adolescents with DLD. Taken together, our findings indicate that French-speaking adolescents with DLD still have deficits in oral language as basic as irregular verb agreement production. These deficits should be addressed in SLP intervention, and to a greater extent in regular classroom settings. Although instruction occurs primarily through written language in high school, it is essential that intervention for adolescents with DLD continue to target oral language, as this remains the source of their difficulties whether in the oral or written form.

2.6. References

Akoglu, H. (2018). User's guide to correlation coefficients. Turkish Journal of Emergency Medicine, 18(3), 91-93. https://doi.org/10/ggw2tg

Abdi, H. (2010). Holm's Sequential Bonferroni Procedure. 8.

Akoglu, H. (2018). User's guide to correlation coefficients. Turkish Journal of Emergency Medicine, 18(3), 91–93. https://doi.org/10/ggw2tg

Altmann, A., Toloşi, L., Sander, O., & Lengauer, T. (2010). Permutation importance: A corrected feature importance measure. Bioinformatics, 26(10), 1340–1347. https://doi.org/10/cm7h6d

American Psychiatric Association. (2013). Neurodevelopmental Disorders. In Diagnostic and statistical manual of mental disorders DSM-5. (5th ed., pp. 31–86). American Psychiatric Publishing. https://doi.org/10.1176/appi.books.9781585624836.jb01

Arslan, S., Broc, L., Olive, T., & Mathy, F. (2020). Reduced deficits observed in children and adolescents with developmental language disorder using proper nonverbalizable span tasks. Research in Developmental Disabilities, 96, 103522. https://doi.org/10.1016/j.ridd.2019.103522

Azhagusundari, B., & Thanamani, A. S. (2013). Feature selection based on information gain. International Journal of Innovative Technology and Exploring Engineering (IJITEE), 2(2), 18–21.

Baddeley, A. (2000). The episodic buffer: A new component of working memory? Trends in Cognitive Sciences, 4(11), 417–423. https://doi.org/10.1016/S1364-6613(00)01538-2

Bishop, D. V. M., Snowling, M. J., Thompson, P. A., & Greenhalgh, T. (2017). Phase 2 of CATALISE: A multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. Journal of Child Psychology and Psychiatry, 58(10), 1068–1080. https://doi.org/10.1111/jcpp.12721

Bouchard, M. G., Fitzpatrick, E. M., & Olds J. (2009). Psychometric analysis of assessment tools used with francophone children. Canadian Journal of Speech-Language Pathology & Audiology, 33(3), 129–139.

Bourget, M.-J. (1987). Variation phonétique dans l'emploi des pronoms de troisième personne en français. In J. Auger, R. Grenier, R. Lapalme, J.-F. Montreuil, & P. Whitmore (Eds.), Tendances actuelles de la recherche sur la langue parlee [Current Trends in Research on Oral Language]: Vol. B-166 (pp. 129–140). Université Laval.

Breault, C., Béliveau, M.-J., Labelle, F., Valade, F., & Trudeau, N. (2019). Le trouble développemental du langage (TDL): Mise à jour interdisciplinaire. Neuropsychologie Clinique et Appliquée, 3(automne 2019), 64–81. https://doi.org/10.46278/j.ncacn.20190717

Brunner, E., & Munzel, U. (2000). The Nonparametric Behrens-Fisher Problem: Asymptotic Theory and a Small-Sample Approximation. Biometrical Journal, 42(1), 17–25. https://doi.org/10.1002/(SICI)1521-4036(200001)42:1<17::AID-BIMJ17>3.0.CO;2-U

Conti-Ramsden, G. (2008). Heterogeneity of specific language impairment in adolescent outcomes. In Understanding developmental language disorders: From theory to practice (pp. 115–129). Psychology Press.

Conti-Ramsden, G., Botting, N., & Faragher, B. (2001). Psycholinguistic markers for specific language impairment (SLI). Journal of Child Psychology and Psychiatry, 42(6), 741–748. https://doi.org/10.1111/1469-7610.00770

Corsi, P. M. (1972). Human memory and the medial temporal region of the brain (unpublished doctoral thesis). McGill University, Montreal.

Courteau, É., Martignetti, L., Royle, P., & Steinhauer, K. (2019). Eliciting ERP components for morphosyntactic agreement mismatches in perfectly grammatical sentences. Frontiers in Psychology, 10. https://doi.org/10.3389/fpsyg.2019.01152

Courteau, É., Royle, P., Gascon, A., Marquis, A., Drury, J. E., & Steinhauer, K. (2013). Gender concord and semantics processing in french children: An auditory ERP study. In S. Baiz, N. Goldman, & R. Hawkes (Eds.), Proceedings of the 37th annual BUCLD (Vol. 1, pp. 87–99).

Courteau, É., Royle, P., & Steinhauer, K. (in preparation). Sentence processing by French teenagers with and without developmental language disorder.

Daniel, T. A., Katz, J. S., & Robinson, J. L. (2016). Delayed match-to-sample in working memory: A BrainMap meta-analysis. Biological Psychology, 120, 10–20. https://doi.org/10.1016/j.biopsycho.2016.07.015

Denman, D., Speyer, R., Munro, N., Pearce, W. M., Chen, Y.-W., & Cordier, R. (2017). Psychometric Properties of Language Assessments for Children Aged 4–12 Years: A

Systematic Review. Frontiers in Psychology, 8, 1515. https://doi.org/10.3389/fpsyg.2017.01515

Dunn, L., Thériault-Whalen, C., & Dunn, L. (1993). Échelle de vocabulaire en images Peabody: Adaptation française du Peabody Picture Vocabulary Test (PsyCan).

Duquette, A.-S., Courteau, E., & Royle, P. (2020, January 31). Fidélité inter-juge d'une étude sur le trouble développemental du langage et de la compréhension du langage oral. 53e Congrès Premier des stagiaires de recherche du 1er cycle de la Faculté de médecine, Montréal, QC. https://doi.org/10.13140/RG.2.2.27265.58722

Ebbels, S. H., Dockrell, J. E., & Lely, H. K. J. van der. (2012). Non-word repetition in adolescents with specific language impairment (SLI). International Journal of Language & Communication Disorders, 47(3), 257–273. https://doi.org/10.1111/j.1460-6984.2011.00099.x

Elin Thordardottir, Kehayia, E., Mazer, B., Lessard, N., Majnemer, A., Sutton, A., Trudeau, N., & Chilingaryan, G. (2011). Sensitivity and specificity of French language and processing measures for the identification of primary language impairment at age 5. Journal of Speech, Language, and Hearing Research, 54(2), 580–597. https://doi.org/10.1044/1092-4388(2010/09-0196)

Elin Thordardottir, & Namazi, M. (2007). Specific language impairment in French-speaking children: Beyond grammatical morphology. Journal of Speech, Language, and Hearing Research, 50(3), 698–715. https://doi.org/1092-4388/07/5003-0698

Franck, J., Cronel-Ohayon, S., Chillier, L., Frauenfelder, U. H., Hamann, C., Rizzi, L., & Zesiger, P. (2004). Normal and pathological development of subject–verb agreement in speech production: A study on French children. Journal of Neurolinguistics, 17(2), 147–180. https://doi.org/10.1016/S0911-6044(03)00057-5

Godard, L., & Labelle, M. (1995). Utilisation de l'ÉVIP avec une population Québécoise (Use of the EVIP with a Quebec population). Fréquences, 7, 18–21.

Greenwell, B., M., & Boehmke, B., C. (2020). Variable Importance Plots—An Introduction to the vip Package. The R Journal, 12(1), 343. https://doi.org/10/gjpj9s

Haebig, E., Weber, C., Leonard, L. B., Deevy, P., & Tomblin, J. B. (2017). Neural patterns elicited by sentence processing uniquely characterize typical development, SLI recovery, and SLI persistence. Journal of Neurodevelopmental Disorders, 9(1), 22. https://doi.org/10.1186/s11689-017-9201-1

Hajian-Tilaki, K. (2013). Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation. Caspian Journal of Internal Medicine, 4(2), 627–635.

Hastie, T., Qian, J., & Tay, K. (2016). An Introduction to glmnet.

Hui, W., Gel, Y. R., Gastwirth, J. L., & Miao, W. (2020). Brunnermunzel.

Jaeschke, R., Guyatt, G. H., Sackett, D. L., Guyatt, G., Bass, E., Brill-Edwards, P., Browman, G., Cook, D., Farkouh, M., & Gerstein, H. (1994). Users' guides to the medical literature: III. How to use an article about a diagnostic test B. What are the results and will they help me in caring for my patients? Jama, 271(9), 703–707.

Justice, L. M., Lomax, R., O'Connell, A., Pentimonti, J., Petrill, S. A., Piasta, S. B., Gray, S., Restrepo, M. A., Cain, K., & Catts, H. (2015). The dimensionality of language ability in young children. Child Development, 86(6), 1948–1965. https://doi.org/10.1111/cdev.12450

Krippendorff, K. (2004). Reliability in Content Analysis. Human Communication Research, 30(3), 411–433. https://doi.org/10.1111/j.1468-2958.2004.tb00738.x

Kuhn, M. (2009). The caret package. Journal of Statistical Software, 28(5).

Leclercq, A.-L., Quémart, P., Magis, D., & Maillart, C. (2014). The sentence repetition task: A powerful diagnostic tool for French children with specific language impairment. Research in Developmental Disabilities, 35(12), 3423–3430. https://doi.org/10.1016/j.ridd.2014.08.026

Leonard, L. B. (2014). Specific language impairment across languages. Child Development Perspectives, 8(1), 1–5. https://doi.org/10.1111/cdep.12053

Leonard, L. B., Miller, C., & Gerber, E. (1999). Grammatical morphology and the lexicon in children with specific language impairment. Journal of Speech, Language, and Hearing Research, 42(3), 678–689.

Lonigan, C. J., & Milburn, T. F. (2017). Identifying the Dimensionality of Oral Language Skills of Children With Typical Development in Preschool Through Fifth Grade. Journal of Speech, Language, and Hearing Research, 60(8), 2185–2198. https://doi.org/10.1044/2017

Maillart, C., & Schelstraete, M.-A. (2005). Grammaticality judgment in French-speaking children with specific language impairment. Journal of Multilingual Communication Disorders, 3(2), 103–109. https://doi.org/10.1080/14769670500066479

McGregor, K. K., Oleson, J., Bahnsen, A., & Duff, D. (2013). Children with developmental language impairment have vocabulary deficits characterized by limited breadth and depth. International Journal of Language & Communication Disorders, 48(3), 307–319. https://doi.org/10.1111/1460-6984.12008

Miller, C. A., Leonard, L. B., & Finneran, D. (2008). Grammaticality judgements in adolescents with and without language impairment. International Journal of Language & Communication Disorders, 43(3), 346–360. https://doi.org/10.1080/13682820701546813

Monetta, L., Desmarais, C., MacLeod, A. A., St-Pierre, M.-C., Bourgeois-Marcotte, J., & Perron, M. (2016). Recension des outils franco-québécois pour l'évaluation des troubles du langage et de la parole. Canadian Journal of Speech-Language Pathology & Audiology, 40(2).

Murtagh, F., & Legendre, P. (2014). Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion? *Journal of Classification*, *31*(3), 274–295. https://doi.org/10.1007/s00357-014-9161-z

Noonan, N. B., Redmond, S. M., & Archibald, L. M. D. (2014). Contributions of Children's Linguistic and Working Memory Proficiencies to Their Judgments of Grammaticality. Journal of Speech, Language, and Hearing Research, 57(3), 979–989. https://doi.org/10.1044/2014_JSLHR-L-12-0225

Plante, E., & Vance, R. (1994). Selection of Preschool Language Tests. Language, Speech, and Hearing Services in Schools, 25(1), 15–24. https://doi.org/10.1044/0161-1461.2501.15

Poulin, M.-J., Marquis, A., & Royle, P. (2015). Étude de faisabilité portant sur l'évaluation de la production et de la compréhension du langage oral en français. In M. Pomerleau & E. M. Gendron-Pontbrian (Eds.), ScriptUM: la revue du colloque VocUM (Vol. 1, pp. 54–68).

Pourquié, M., Courteau, É., Duquette, A.-S., & Royle, P. (in revision). Verb inflection and argument structure processing in French adolescents with DLD. Journal of Child Language

Prigent, G., Parisse, C., Leclercq, A.-L., & Maillart, C. (2015). Complexity markers in morphosyntactic productions in French-speaking children with specific language impairment (SLI). Clinical Linguistics & Phonetics, 29(8–10), 701–718. https://doi.org/10.3109/02699206.2015.1020451

Quémart, P., & Maillart, C. (2016). The sensitivity of children with SLI to phonotactic probabilities during lexical access. Journal of Communication Disorders, 61, 48–59. https://doi.org/10.1016/j.jcomdis.2016.03.005

Redmond Sean M., Ash Andrea C., & Hogan Tiffany P. (2015). Consequences of Co-Occurring Attention-Deficit/Hyperactivity Disorder on Children's Language Impairments. Language, Speech, and Hearing Services in Schools, 46(2), 68–80. https://doi.org/10.1044/2014_LSHSS-14-0045

Rice, M. L., & Hoffman, L. (2015). Predicting Vocabulary Growth in Children With and Without Specific Language Impairment: A Longitudinal Study From 2;6 to 21 Years of Age. Journal of Speech, Language, and Hearing Research, 58(2), 345–359. https://doi.org/10.1044/2015_JSLHR-L-14-0150

Rietveld, T., & van Hout, R. (2015). The t test and beyond: Recommendations for testing the central tendencies of two independent samples in research on speech, language and hearing pathology. Journal of Communication Disorders, 58, 158–168. https://doi.org/10.1016/j.jcomdis.2015.08.002

Rose, Y., & Royle, P. (1999). Uninflected structure in familial language impairment: Evidence from French. Folia Phoniatrica et Logopaedica, 51(1–2), 70–90. https://doi.org/10.1159/000021482

Royle, P., Mazzocca, P., St-Denis, A. et Marquis, A. (2018). Insensitivity to verb conjugation patterns in French children with SLI. Clinical Linguistics & Phonetics, 32(2), 128–147. doi: 10.1080/02699206.2017.1328706

RStudio Team. (2020). RStudio: Integrated Development for R. RStudio, PBC; http://www.rstudio.com/.

Secord, W. A., Wiig, E., Boulianne, L., Semel, E., & Labelle, M. (2009). Évaluation clinique des notions langagières fondamentales®—Version pour francophones du Canada (CELF® CDN-F). The Psychological Corporation.

Statistics Canada. (2016). Mother tongue for the total population excluding institutional residents (100% data), Census Profile Quebec and Canada. https://www12.statcan.gc.ca/

Tessier, A., & Valade, S. (2017). Organisation du continuum et de la dispensation des services aux enfants âgés de 2 à 9 ans présentant un trouble développemental du langage (trouble primaire du langage) (p. 107). Gouvernement du Quebec. http://collections.banq.qc.ca/ark:/52327/3185271

Thiele, C., & Hirschfeld, G. (2020). Cutpointr: Improved estimation and validation of optimal cutpoints in R. ArXiv Preprint ArXiv:2002.09209.

Tomblin, J. B., & Zhang, X. (2006). The dimensionality of language ability in school-age children. Journal of Speech, Language, and Hearing Research, 49(6), 1193–1208. https://doi.org/10.1044/1092-4388(2006/086)

Tuller, L., Henry, C., Sizaret, E., & Barthez, M.-A. (2012). Specific language impairment at adolescence: Avoiding complexity. Applied Psycholinguistics, 33(1), 161–184. https://doi.org/10.1017/S0142716411000312

Ullman, M. T., & Pierpont, E. I. (2005). Specific language impairment is not specific to language: The procedural deficit hypothesis. Cortex, 41(3), 399–433. https://doi.org/10.1016/S0010-9452(08)70276-4

Van Kleeck, A. (1982). The emergence of linguistic awareness: A cognitive framework. Merrill-Palmer Quarterly, 28(2), 237–265.

Zhang, J., & Mueller, S. T. (2005). A note on ROC analysis and non-parametric estimate of sensitivity. Psychometrika, 70(1), 1–10. https://doi.org/10.1007/s11336-003-1119-8

2.7. Supplementary materials manuscript 1

2.7.1. Detailed tasks presentation

The tasks used in this study can be classified into two categories and will be briefly described in the following order: 1) those commonly used by speech-language pathologists (SLP) in clinical settings to assess language skills of Québec French adolescents, and 2) those used in research on developmental language disorder (DLD).

Three tasks were selected from the CELF-IV$^{cnd\text{-}F}$ French version standardized among Québec-French speakers ages 4 to 16 (Secord et al., 2009) and were administered as recommended by the manual. The Recalling Sentences task, where participants needed to repeat orally-presented sentences without any word changes, assessed lexico-semantics, morphosyntax and phonological working memory skills (Leclercq et al., 2014). The Word Classes task assessed ability to understand the lexico-semantic relationships between orally presented words by choosing two words that go together in a choice of four (receptive subtask) and to explain this relationship (expressive subtask). The number repetition tasks consisted in a forward and backward digit span. We also used the non-standardized French Québec Nonword repetition Courcy task (Elin Thordardottir et al., 2011) which consists of 40 words ranging in length from two to five syllables. Scoring followed the task's recommendation: phoneme omissions and substitutions were counted as incorrect, while distortions and additions did not result in point loss. A point was given for each repeated phoneme, with a maximum of 280, and we used the total repeated phonemes as participants' score. Neurotypical adults repeat between 277 and 279 phonemes correctly, as revealed by a task pre-validation with 10 French-speakers having French as their daily language (Duquette et al., 2020). We chose the EVIP task, a standardized Canadian French version of the Peabody Vocabulary Test for 2:5 to 18 years old (Dunn et al., 1993) to evaluate the receptive vocabulary. In this, participants choose among four pictures the one matching a word spoken by the experimenter. Expressive vocabulary was assessed with an action (verb) naming task taken from the French version of the fLEX test (task 2 see Pourquié et al, in revision), where participants had to describe with a verb each of 30 actions depicted on pictures. Subject-verb number agreement production and comprehension skills were assessed through tasks three and four of the fLEX test (*ibid*), which each contained 35 items. To target only irregular verbs that had an audible agreement number cue on the verb's ending (e.g. *il rugit* [ilʁyʒi], 'he roars' vs. *ils rugissent* [ilʁyʒɪs], 'they roar'), we rated a subset of 20 items from the original tasks, bringing the maximum score to 20. The expressive task assessed sentence production of inflected verbs in the present tense, either in the singular or plural depending on the number of agents depicted on the picture. The receptive task assessed understanding of inflected verbs in the singular and plural: the participant chose among four pictures the one that matched a sentence spoken by the experimenter. This subtask used a sentence-picture matching paradigm with one target image and three foils (a number agreement-error, e.g., one lion roaring for *they roar*, a lexical error on the verb, e.g., one lion sleeping, or both, e.g., two lions sleeping).

We used two grammaticality judgment tasks where an alien comes to Québec to learn French and sometimes makes mistakes (Courteau et al., 2013). These data were taken from an off-line grammaticality judgment task and an event-related potentials (ERP) experimental session (Courteau et al., in preparation). During the off-line task, participants had to listen to pre-recorded sentences while looking at pictures and judge if sentences were correct or not by answering yes or no. The first task was adapted from Poulin et al. (2015). Participants listened to 16 sentences while watching pictures that were either correct (4) or contained errors targeting the noun phrase (n = 12). Errors included auditory-visual lexico-semantic mismatches on nouns (e.g., visual [BROWN SHOE ON TABLE], *Je vois un !train brun...*, ‘I see a brown !train…’, *n* = 4) and morphosyntactic gender-agreement errors on the determiner (e.g., *Je vois *la soulier vert ...,* ‘I see *the.F shoe.M green.M…’, *n* = 4) or the adjective (e.g., *Je vois le soulier *verte ...,* ‘I see the.M shoe.M *green.F…’, *n* = 4). In the second task run during the EEG recording, participants listened to 300 sentences while watching pictures that were either a match (150) or contained errors that targeted the verb (150) and judged if the visuo-auditory pairs were a match or not using a button press. Lexico-semantic errors were created with a verb that did not match the depicted action (*n* = 30, e.g., visual [A WOMAN SINGS], ... *elle !nage dans la piscine publique,* ‘she !swims in the public pool’). Subject-verb number agreement errors were created by varying the number of visually presented agents and morphosyntactic number cues in the auditory stimuli. All auditory cues were perfectly grammatical (Courteau et al, 2019). This was operationalized using either verbs with regular agreement morphophonology, where the plural number cue “s” [z] is created by the liaison between the pronoun plural form and verb’s vowel onset (*n* = 60, e.g., visual [A GIRL EATS], *Au dessert, *elles‿aiment [ɛlzɛm] la mousse au chocolat,* ‘For dessert, *they like chocolate mousse’) or with verbs whose number cue was audible on the verb ending with irregular morphophonology (*n* = 60, e.g., visual [A LION ROARS], *En soirée, ils *rugissent* [ilʁyʒɪs] *dans la savane,* ‘In the evening, *they roar in the jungle’).

Further, we tested participants with commonly used tasks in DLD research assessing nonverbal visual working memory[3]. We used 4 computer-based nonverbal working memory

[3] For the purposes of this article, we will use the term working memory in the sense of “a limited capacity system allowing the temporary storage and manipulation of information’ as defined by Baddeley, (2000, p.418). This system includes a phonological and a visuospatial component, which we will refer as the phonological and visuospatial working memory.

tasks (Cognitive Experiments IV v2 pack of the Presentation® software, Version 18.0, Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com). Of these was the forward and backward Corsi Blocks tasks (Corsi, 1972), where a sequence of highlighted squares is presented on the computer screen, and the participant must recreate the sequence using the mouse in forward or backward order. We used a delayed match-to-sample task of non-verbal stimuli (Daniel et al., 2016) where a form made of sixteen squares is displayed, and after a delay of 1 or 5 seconds, the participant must recall the form by choosing the right one in a choice of two. See supplementary Table 1 for the type of score available (raw, aged-based percentile, *A*-score), and the underlying linguistic or working memory subdomains assessed by each subtask, based on the tests' manual when available, or the literature.

Supplementary Table 1

Complete list of subtasks and the underlying linguistic and cognitive subdomain they assessed

Subtasks	Scores	Linguistic and cognitive subdomains
Subtasks used in clinical settings		
Recalling	Raw, Pcl	Lexico-semantics, Morphology and Syntax (Leclercq et al., 2014)
Word Classes	Raw, Pcl	Lexico-semantic classes' relationship (CELF-IV$^{cnd\text{-}F}$)
Forward Digit	Raw, Pcl	Verbal working memory (CELF-IV$^{cnd\text{-}F}$; Baddeley et al., 2000)
Backward Digit	Raw, Pcl	Verbal working memory (CELF-IV$^{cnd\text{-}F}$; Baddeley et al., 2000)
Nonword rep.	Raw	Verbal working memory (Gathercole et al., 1994)
EVIP	Raw, Pcl	Lexico-semantics: receptive vocabulary (Dunn et al., 1993)
Action naming	Raw	Lexico-semantics: lexical access of verbs (Pourquié et al., in revision)
Irr. verb prod.	Raw	Morphosyntactic processing of irr. verbs (Pourquié et al., in revision)
Irr. verb comp.	Raw	Morphosyntactic processing of irr. verbs (Pourquié et al., in revision)
Subtasks used in research on DLD		
G. j.: Nouns	*A*-score	Metaling.: semantic content, auditory-visual modality (Van Kleeck, 1982)
G. j.: Det.	*A*-score	Metaling.: morphosyntactic content, auditory modality (*ibid*)
G. j.: Adj.	*A*-score	Metaling.: morphosyntactic content, auditory modality (*ibid*)
G. j.: Verbs	*A*-score	Metaling.: semantic content, auditory-visual modality (*ibid*)
G. j.: Regular	*A*-score	Metaling.: morphosyntactic content, auditory-visual modality (*ibid*)
G. j.: Irr. agrm.	*A*-score	Metaling.: morphosyntactic content, auditory-visual modality (*ibid*)
Corsi–Forward	Raw	Visuospatial working memory (Corsi, 1972; Baddeley et al., 2000)
Corsi–Backward	Raw	Visuospatial working memory (Corsi, 1972; Baddeley et al., 2000)
DMTS–1s	Raw	Visuospatial working memory (Daniel et al., 2016)
DMTS–5s	Raw	Visuospatial working memory (Daniel et al., 2016)

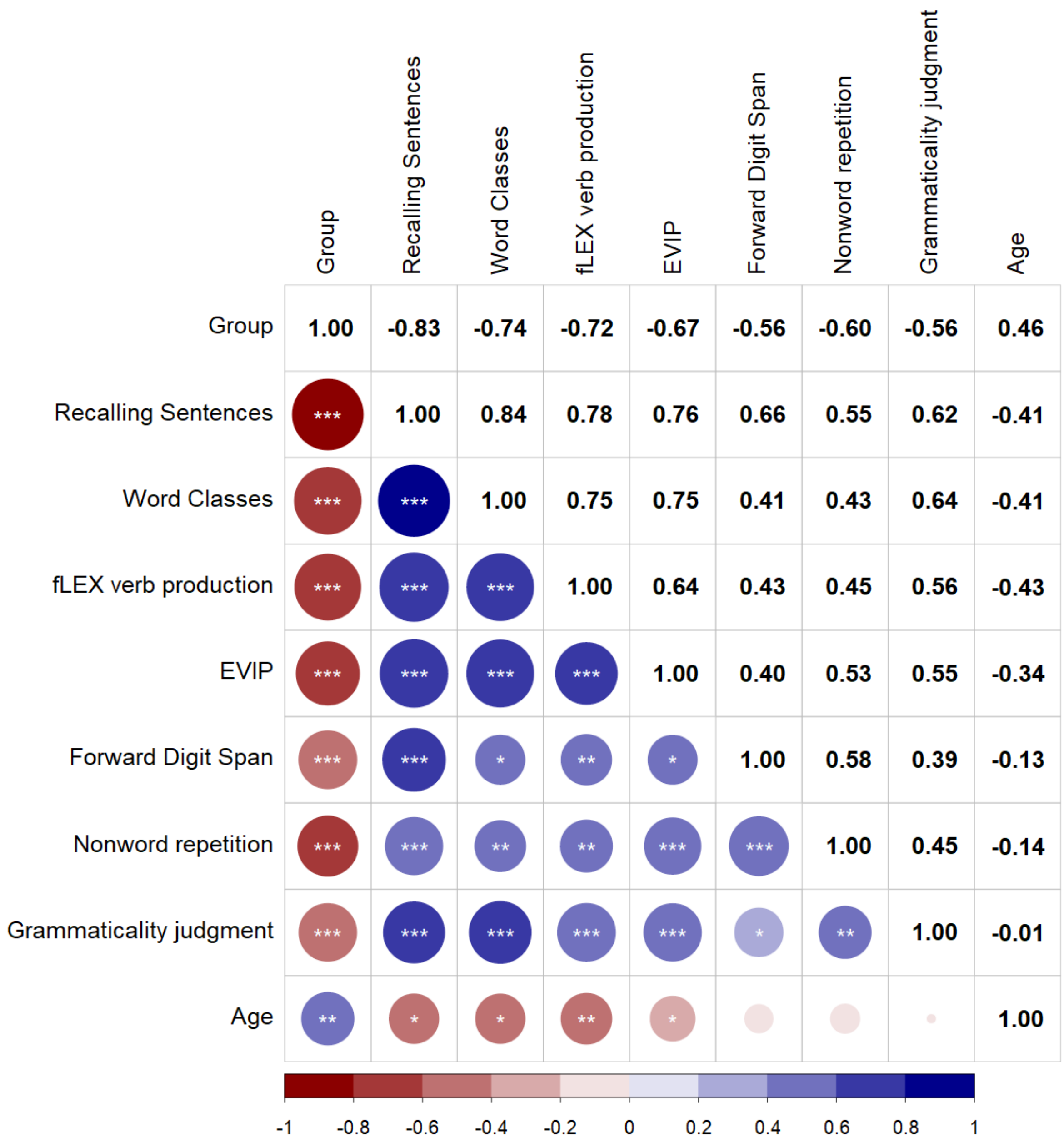
Note. Pcl, Percentile; Nonword rep., Nonword repetition task; Metaling., Metalinguistics; G. j., Irr., irregular; Grammaticality judgment; Det., Determiner; Adj., Adjective; Reg. agrm., Regular subject-verb agreement; Irr. agrm., Irregular subject-verb agreement; DMTS, Delayed Match to Sample.

2.7.2. Correlational analyses

Supplementary Figure 1 shows the correlation matrix and the Spearman correlation statistics plot for Age, Group, and the seven selected subtask scores.

Supplementary Figure 1

Correlation plot with Spearman coefficients for selected subtasks, with group and age.



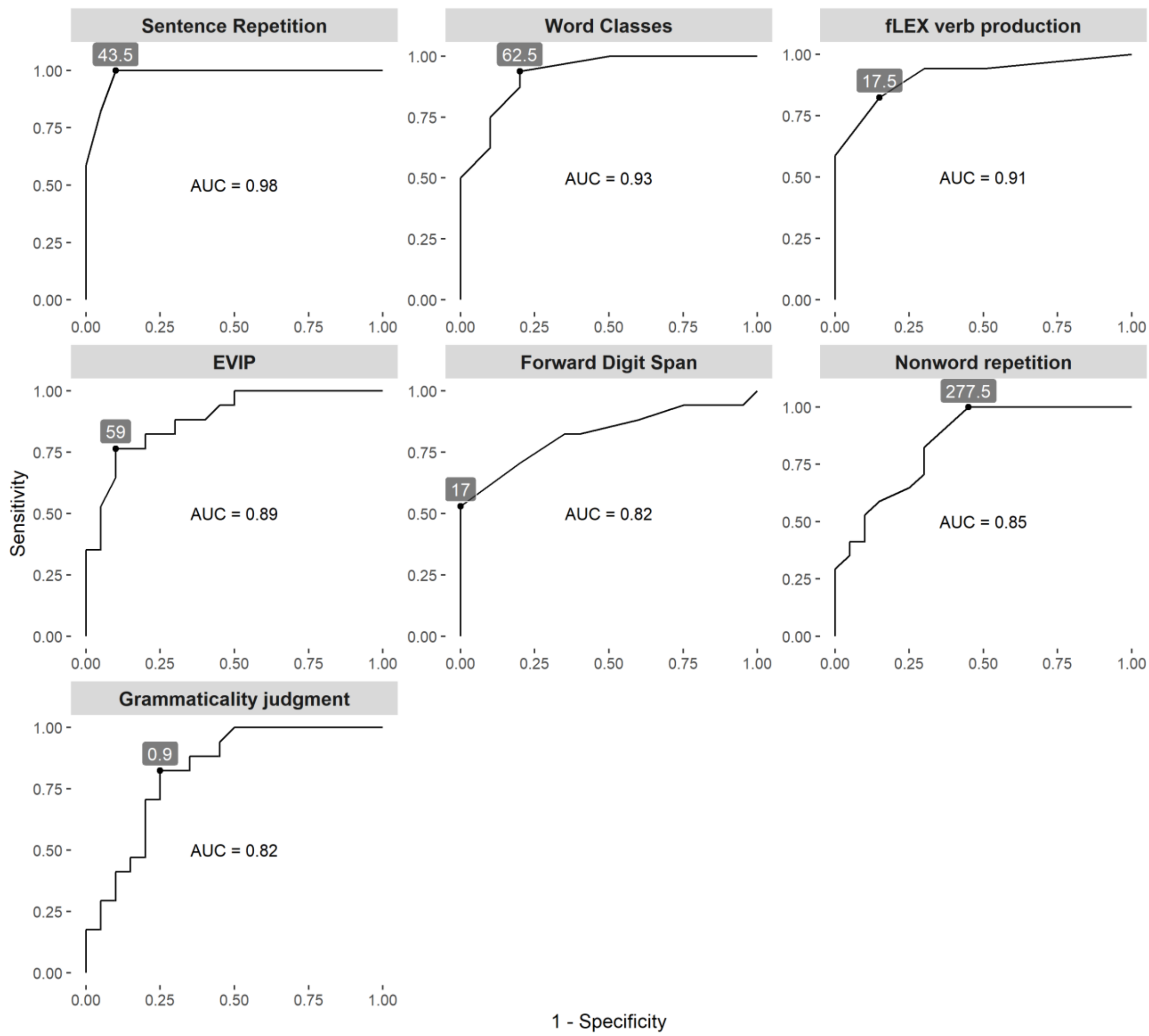
Note. Subtask Grammaticality judgment: Grammaticality judgment of irregular verbs. Coefficients and levels of statistical significance are based on Spearman rank correlations. *p < 0.05, **p < 0.1, ***p < 0.001.

2.7.3. Optimal cut-off scores

We present for the seven discriminating subtasks their ROC plot showing the area under the curve for all possible scores and with their respective optimal cut-off scores.

Supplementary Figure 2

ROC curves for all seven discriminating subtasks and their optimal cut-off scores



Note. Grey labels indicate optimal cut-off scores. Three score types are displayed: percentile scores (Recalling Sentences, Word Classes, Forward Digit Span and EVIP), raw scores (fLEX irregular verb production: maximum of 20 verbs, Nonword repetition: maximum of 280 phonemes) and one *A-score* (Grammaticality judgment of irregular verb agreement).

2.7.4. References

Brunner, E., & Munzel, U. (2000). The Nonparametric Behrens-Fisher Problem: Asymptotic Theory and a Small-Sample Approximation. *Biometrical Journal*, *42*(1), 17–25. https://doi.org/10.1002/(SICI)1521-4036(200001)42:1<17::AID-BIMJ17>3.0.CO;2-U

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 417–423. https://doi.org/10.1016/S1364-6613(00)01538-2

Corsi, P. M. (1972). *Human memory and the medial temporal region of the brain (unpublished doctoral thesis)*. McGill University, Montreal.

Courteau, É., Royle, P., Gascon, A., Marquis, A., Drury, J. E., & Steinhauer, K. (2013). Gender concord and semantic processing in french children: An auditory ERP study. In S. Baiz, N. Goldman, & R. Hawkes (Eds.), *Proceedings of the 37th annual BUCLD* (Vol. 1, pp. 87–99).

Courteau, É., Royle, P., & Steinhauer, K. (in preparation). *Sentence processing by French teenagers with and without developmental language disorder*.

Daniel, T. A., Katz, J. S., & Robinson, J. L. (2016). Delayed match-to-sample in working memory: A BrainMap meta-analysis. *Biological Psychology*, *120*, 10–20. https://doi.org/10.1016/j.biopsycho.2016.07.015

Dunn, L., Thériault-Whalen, C., & Dunn, L. (1993). *Échelle de vocabulaire en images Peabody: Adaptation française du Peabody Picture Vocabulary Test* (PsyCan).

Duquette, A.-S., Courteau, E., & Royle, P. (2020, January 31). *Fidélité inter-juge d'une étude sur le trouble développemental du langage et de la compréhension du langage oral*. 53e Congrès Premier des stagiaires de recherche du 1er cycle de la Faculté de médecine, Montréal, QC. https://doi.org/10.13140/RG.2.2.27265.58722

Elin Thordardottir, Kehayia, E., Mazer, B., Lessard, N., Majnemer, A., Sutton, A., Trudeau, N., & Chilingaryan, G. (2011). Sensitivity and specificity of French language and processing measures for the identification of primary language impairment at age 5. *Journal of Speech, Language, and Hearing Research*, *54*(2), 580–597. https://doi.org/10.1044/1092-4388(2010/09-0196)

Gathercole, S. E., Willis, C. S., Baddeley, A. D., & Emslie, H. (1994). The children's test of nonword repetition: A test of phonological working memory. *Memory*, *2*(2), 103–127. https://doi.org/10.1080/09658219408258940

Leclercq, A.-L., Quémart, P., Magis, D., & Maillart, C. (2014). The sentence repetition task: A powerful diagnostic tool for French children with specific language impairment. *Research in Developmental Disabilities*, *35*(12), 3423–3430. https://doi.org/10.1016/j.ridd.2014.08.026

Poulin, M.-J., Marquis, A., & Royle, P. (2016). Étude de faisabilité portant sur l'évaluation de la production et de la compréhension du langage oral en français. *ScriptUM: La Revue Du Colloque VocUM*. http://journal.vocum.ca/index.php/scriptum/article/view/28

Pourquié, M., Courteau, É., Duquette, A.-S., & Royle, P. (in revision). Verb inflection and argument structure processing in French adolescents with DLD. *Journal of Child Language*

Secord, W. A., Wiig, E., Boulianne, L., Semel, E., & Labelle, M. (2009). *Évaluation clinique des notions langagières fondamentales®—Version pour francophones du Canada (CELF® CDN-F)*. The Psychological Corporation.

Van Kleeck, A. (1982). The emergence of linguistic awareness: A cognitive framework. *Merrill-Palmer Quarterly*, *28*(2), 237–265.

3. Bridge 1

In this first manuscript, we investigated subregular and irregular (SUBIRR) subject-verb number agreement at the behavioural level with linguistic tasks in (pre-)teenagers with and without DLD. Results suggested, among others, that tasks assessing lexico-semantics, based on the semantic relationships between words, and morphosyntactic production of SUBIRR subject-verb number agreement were the best at discriminating our participants. Furthermore, when we combined our most discriminating tasks, we found that the SUBIRR production task contributed the most to the multivariable model for the identification of participants with DLD. To test and compare morphosyntactic to lexico-semantic processing at the neurocognitive level, we developed an ERP experiment in which these linguistic domains were observable. This experiment needed to be suitable for children and teenagers with and without DLD. Furthermore, we first needed to investigate what ERP components French-speaking adults would elicit in this experiment.

In the second manuscript, we present a novel ERP experiment investigating morphosyntactic and lexico-semantic processing with *only grammatical sentences*. This experiment was administered to 28 French-speaking adults, thus revealing what would be expected in mature processing in our conditions.

4. Manuscript 2

Eliciting ERP components for morphosyntactic agreement mismatches in perfectly grammatical sentences

Émilie Courteau[1,2], Lisa Martignetti[3], Phaedra Royle [1,2] Karsten Steinhauer [2,3]

[1]School of Speech Language Pathology and Audiology, Faculty of Medicine, University of Montreal, Montréal, Quebec, Canada

[2]Centre for Research on Brain, Language and Music (CRBLM), Montreal, Quebec, Canada

[3]School of Communication Sciences and Disorders, Faculty of Medicine, McGill University, Montréal, Quebec, Canada

This manuscript has been published in Frontiers in Psychology, 2019, Volume 10.

https://doi.org/10.3389/fpsyg.2019.01152

Abstract

The present event-related brain potential (ERP) study investigates mechanisms underlying the processing of morphosyntactic information during real-time auditory sentence comprehension in French. Using an auditory-visual sentence-picture matching paradigm, we investigated two types of anomalies using entirely grammatical auditory stimuli: (i) semantic mismatches between visually presented actions and spoken verbs, and (ii) number mismatches between visually presented agents and corresponding morphosyntactic number markers in the spoken sentences (determiners, pronouns in liaison contexts, and verb-final "inflection"). We varied the type and amount of number cues available in each sentence using two manipulations. First, we manipulated the verb type, by using verbs whose number cue was audible through subject (clitic) pronoun liaison (liaison verbs) as well as verbs whose number cue was audible on the verb ending (consonant-final verbs). Second, we manipulated the pre-verbal context: each sentence was preceded either by a neutral context providing no number cue, or by a subject noun phrase containing a subject number cue on the determiner.

Twenty-eight French-speaking adults participated in the experiment. While sentence judgment accuracy was high, participants' ERP responses were modulated by the type of mismatch encountered. Lexico-semantic mismatches on the verb elicited the expected N400 and additional negativities. Determiner number mismatches elicited early anterior negativities, N400s and P600s. Verb number mismatches elicited biphasic N400-P600 patterns. However, pronoun+verb liaison mismatches yielded this pattern only in the plural, while consonant-final changes did so in the singular and the plural. Furthermore, an additional sustained frontal negativity was observed in two of the four verb mismatch conditions: plural liaison and singular consonant-final forms.

This study highlights the different contributions of number cues in oral language processing and is the first to investigate whether auditory-visual mismatches can elicit errors reminiscent of outright grammatical errors. Our results emphasize that neurocognitive mechanisms underlying number agreement in French are modulated by the type of cue that is used to identify auditory-visual mismatches.

Keywords: subject-verb number agreement, event-related brain potentials (ERPs), auditory-visual sentence-picture matching paradigm, cross-modal number mismatches, French language, online grammaticality judgment, N400 and P600, sustained frontal negativity

4.1. Introduction

Few ERP studies have investigated real-time auditory sentence comprehension in French. French subject-verb agreement has specific properties e.g., clitic-verb liaison and verb-final consonants in two of its three verb groups (see below) relevant to the study of agreement processing, and which have not been systematically studied in the ERP literature. Furthermore, many studies of agreement rely on visual word presentation, where morphosyntactic information is presented simultaneously with other lexical information, rather than unfolding over time, as in natural spoken language. These reading studies may not capture temporal aspects typical of spoken language processing, and ERP components may differ across modalities. Moreover, there is increasing interest in ERP methods that do not rely on violation paradigms. Considering these issues, we developed an ERP study where we implemented an auditory-visual sentence-picture matching task to investigate on-line processing of lexico-semantic and morphosyntactic information. Creating mismatches between grammatical auditory sentences and picture stimuli has been shown to elicit ERPs in lexico-semantic noun mismatches (e.g., Willems et al, 2008). To our knowledge, these mismatches between modalities have not been used to study morphosyntactic processing, nor lexico-semantic verb mismatches. Therefore, we examined whether the auditory presentation of a grammatical sentence combined with a picture that doesn't match its morphosyntactic features would elicit the same ERP components as in classic paradigms that use ungrammatical sentences. Our innovative approach is motivated by the long-term aim of our research program, which is to study language processing in children with developmental language disorder (previously referred to as specific language impairment, SLI) using ecologically valid stimuli. Combining images and speech resembles other common activities such as shared picture-book reading, or watching documentary or educational videos, where an image is presented concurrently with an oral description. In these cases, people being read or spoken to might make predictions about what the reader will say, and notice any incongruencies, as they were expected to do during our experiment. Thus, we investigate: (i) lexical-semantic mismatches between visually presented actions and spoken verbs, and (ii) auditory-visual subject number mismatches while varying number-cue types at different positions in the sentence. These manipulations should allow us to better understand how French-speakers handle semantic and grammatical cues online while processing language, and should also elucidate if cross-modal paradigms elicit similar ERP components as classic within-sentence agreement violations. We will first review relevant ERP findings and then develop our research questions.

In ERPs, lexical-semantic processing is typically reflected by the centro-parietal N400 component between 300-500 ms after word onset (Kutas and Hillyard, 1980). This brain wave can be elicited by lexical-semantic expectancy violations (Kutas and Federmeier, 2000; Steinhauer and Connolly, 2008). Its amplitude may reflect processing effort during lexical retrieval (Lau et al., 2008) and post-lexical integration (Steinhauer et al., 2017), or it can be described as an error signal reflecting the difference between one's lexical-semantic expectations (i.e., the "current model") and the actual word input (Bornkessel-Schlesewsky & Schlesewsky, 2018; henceforth BSS2018). Although most evidence for N400s has come from reading studies, this component has also been observed in bimodal (auditory-visual) lexical semantic violations where an incongruous image is presented concurrently with an auditory utterance, for instance: *Je vois un !<u>soulier</u> vert sur la table* 'I see a green <u>shoe</u> on the table' with an image of a <u>HAT on a table</u> (Royle et al., 2013; see also Friedrich and Friederici, 2004; Willems et al., 2008). The N400 is generally considered a reliable ERP correlate of increased lexico-semantic processing difficulties.

Morphosyntactic agreement-error processing in reading studies is often indexed by one or two components, the left anterior negativity (LAN) and a later positive shift (the P600). The LAN has been reported for a range of morphosyntactic violations, including verb agreement violations (e.g., *As a turtle grows, its shell *grow too* (Kutas and Hillyard, 1983), especially in languages with relatively free word order and rich morphological agreement marking (Angrilli et al., 2002; Barber and Carreiras, 2005), but also in languages with less rich paradigms (Hagoort and Brown, 2000; Osterhout and Mobley, 1995). Like the N400, this component typically emerges between 300 and 500 ms after stimulus presentation. Most agreement studies eliciting LANs have been conducted in the written modality, but some auditory studies have also reported LAN-like negativities for a range of morpho-syntactic anomalies (Friederici et al., 1993; Balconi and Pozzoli, 2005; Rossi et al., 2006; Hasting and Kotz, 2008; Morgan-Short et al., 2010; Dube et al., 2016; Haebig et al., 2017). Compared to reading studies, LANs in auditory studies tend to have an earlier onset, a much longer duration (~100-1200 ms), and a bilateral frontal distribution (e.g., Hasting & Kotz, 2008). However, several reading studies do not report LANs for agreement violations (Osterhout and Mobley, 1995; Lau, Stroud, Plesch, and Phillips, 2006; Tokowicz and MacWhinney, 2005; Nevins et al., 2007; Foucart and Frenck-Mestre, 2011, 2012) and report only P600s (see below). Whether or not LANs are reliable reflections of morphosyntactic processes, whether different morphologies reflect distinct processes, and what their functional significance may be, is therefore under debate (Tanner, 2015; Molinaro et al., 2011; Royle et al., 2013; Steinhauer and Drury, 2012).

The LAN is usually followed by a late parietal positive-going component, the P600, roughly between 500 and 1000 ms (Hahne and Friederici, 1999; Osterhout and Holcomb, 1992, 1993; Steinhauer et al., 1999). In contrast to the LAN, the P600 is widely viewed as the most consistent ERP signature for a large range of grammatical anomalies. It has been observed for gender-agreement and verb-agreement violations (Frenck-Mestre et al., 2008; Foucart and Frenck-Mestre, 2011, 2012; Molinaro et al., 2011a; Royle et al., 2013), syntactic violations (Friederici, 2002), garden path sentences (Osterhout and Holcomb, 1992, 1993), and has also been elicited by semantic anomalies in conjunction with N400s (Hagoort, 2003; Royle et al., 2013; Steinhauer et al., 2010). While many agree that the P600 is a brain response related to controlled sentence reanalysis and repair (Hahne and Friederici, 1999), some argue that it is an ERP correlate of implicit syntactic processing (Tokowicz and MacWhinney, 2005). Another interpretation is that the P600 is a member of the parietal P300 (P3b) family of components reflecting stimulus categorization (e.g., in an acceptability judgment task) (Royle et al., 2013; Sassenhagen et al., 2014; BBS2018).

ERP studies have also revealed different patterns for various agreement error types. A majority of studies on agreement are reading tasks, and most use serial word-by-word visual presentation. Molinaro et al. (2011) present a review of number and gender agreement processing in various languages. Regarding subject-verb number agreement violations, of 17 studies reviewed, all revealed P600s and 13 revealed LANs. The authors correlate the LAN with morphosyntactic error processing and explain the absence of a LAN in certain studies by differences in morphosyntactic saliency. For example, when these are underspecified (i.e., not morphologically expressed on the singular), a LAN may not be triggered. Molinaro et al. (2011b), found that in conditions such as **Il ragazzo e la ragazza corre…* ‘The boy and the girl run.3rd.SINGULAR’, the conjoined noun phrase (NP) does not contain any overt plural marking and in its absence no LAN is triggered. However, Frenck-Mestre et al. (2008) do not observe any negativities resembling a LAN but find a P600 in French native speakers in response to subject-verb agreement violations such as **Le matin je mangez* [mãʒe] ‘In the morning I eat.2nd.PLURAL’. Their data contradict Molinaro et al.'s (2011a) interpretation, as the LAN was absent even though subject number properties were clearly expressed by the singular pronoun *je* ‘I’ as well as the verb *mangez*.

In sum, while both the P600 and the LAN can be observed following various agreement-error types, it is still unclear whether they are modulated by the languages, structures, or contexts used to elicit them. The present study attempts to answer the following questions, using entirely grammatical sentences in all conditions. First, whether French speakers will elicit an N400 component for cross-modal (audio-visual) lexico-semantic mismatches realized on actions/verbs – rather than nouns/objects – and whether this violation type will elicit P600s as observed in other cross-modal lexico-semantic mismatch studies. Second, whether cross-modal “morphosyntactic” number mismatches between the picture’s agents and the determiners/pronouns or verb morphology in our sentences elicit biphasic LAN/N400-P600 complexes as in previous morphosyntactic violation studies. To the best of our knowledge, this has not been investigated before. Given that our sentences were grammatical, one could argue that cross-modal number mismatches may cause either (a) conceptual-semantic problems typically associated with N400s instead of LANs, or (b) logical-semantic conflicts related to truth values, which have be found to elicit local N400s or sentence wrap-up effects (Bokhari, 2015) and P600s followed by (but not preceded by) late LANs (L-LANs; cf. Steinhauer et al., 2010). The third question was whether participants, when presented with multiple cues for number mismatch disambiguation, will rely on the first available auditory cue, as indicated by ERP responses.

4.2. Materials and Methods

4.2.1. Participants

Twenty-eight neurotypical adults aged 18–40 years participated in the experiment. The protocol was approved by Institutional Review Boards at McGill and University of Montreal (UdeM). All participants gave written informed consent in accordance with the Declaration of Helsinki. All were right-handed as assessed using the Edinburgh Handedness Inventory (Oldfield, 1971), had normal or corrected-to-normal vision, French as their mother tongue and their everyday language, and did not learn any other language before age 5. None had learning disabilities, neurological damage, or hearing loss. Working memory was assessed orally at session's end. Participants were recruited from Montreal university student populations. Participants were compensated $45 for their time (3.5 hours). Six data sets had to be excluded due to excessive eye movement artefacts, such that data from 22 participants were retained for analyses (range: 18-38 years; mean 25; 12 female, 10 male). We consider this sample size as enough to provide a good estimate of the effects of interest, since in Royle et al (2013) a group of 15 French-speaking adults participating in a similar paradigm (7 in a task-based group and 8 in a no-task one) showed significant ERPs related to adjective agreement errors and noun-image semantic incongruencies in each group.

4.2.2. Materials and design

As illustrated in Tables 1-3, materials consisted of spoken grammatical sentences in French, half of which mismatched with a concurrently-displayed picture, either through the action described or the number of agents (singular/plural mismatch). As we developed the study for younger populations (to be tested after adults), word selection was constrained by age-of-acquisition norms (see Supplementary Materials for details). Verbs were presented within sentences containing third person singular or plural subject pronouns (*he/she/they*), and a sentence continuation with a direct object NP, or prepositional phrase (PP, e.g., … *in the public pool*) to avoid sentence-final (or "wrap-up") effects in ERPs time-locked to verbs (Hagoort, 2003; see also Stowe et al., 2018). Verbs were selected based on their number agreement morphological characteristics, as explained below.

Selected critical verbs were selected inspired by the fLEX evaluation tool (Pourquié et al., 2016), with their imageability in mind, as they were presented alongside illustrations, and were matched on lemma frequency, age of emergence, and length (syllables and phonemes). Auditory stimulus recording, normalizing and splicing was supervised by trained research assistants with a background in speech editing (Supplementary Materials). For each sentence, one colour drawing was created by a professional artist, emphasizing the action being described, and the agent(s) carrying it out. Drawings maintained a constant visual complexity level, avoiding superfluous or distracting details.

In order to enhance the comparability of ERP effects between semantic and number mismatches, we decided to create semantic mismatches on the verb, the main element disambiguating mismatches in our number conditions (see below). Thus, for semantic mismatches, the spoken verb did not correspond to the depicted action (e.g., the sound file described 'she swims…' and the image depicted 'she sings…'). Sentences in this condition were created with 60 invariable regular verbs, 30 with a singular and 30 with a plural pronoun ('he/she', 'they'). Each pronoun+verb item was then combined with (a) a subject NP context providing a lexical NP with early number information (e.g., ‘The.PLURAL girls, they swim’[4] and (b) a neutral context without number information (e.g., ‘In the evening, they swim’), resulting in 120 spoken items. In total, 300 stimuli were created; 150 congruent and 150 in incongruent ones, by splicing the incongruent verb into the sentence (see e.g., Table 1. 2a).

Table 1

Experimental sub-conditions for lexico-semantic manipulations and a corresponding visual stimulus.

Visual Stimulus

Sample visual stimulus presented concurrently with auditory stimuli for matching lexico-semantic conditions (1a-b) and mismatching ones (2a-b). Note that, in addition to the mismatch at the target verb (“sings” vs. “swims”), conditions 2a-b also include a second semantic mismatch in the prepositional phrase (here: “concert venue” vs. “public pool”).

Condition	**Context**	**Sample auditory stimuli**	
Semantic Congruent	Neutral	(1a)	*Chaque semaine ǀ elle chante dans une salle d concert* ‘Each week ǀ she sings at a concert venue’
	Subject NP	(1b)	*La vedette ǀ elle chante dans une salle de concert* ‘The star ǀ she sings at a concert venue’
Semantic Incongruent	Neutral	(2a)	*Chaque semaine ǀ elle !nage dans la piscine publique* ‘Each week ǀ she !swims in the public pool’
	Subject NP	(2b)	*La vedette ǀ elle !nage dans la piscine publique* ‘The star ǀ she !swims in the public pool’

Note. Critical words are underlined. Subj NP = overt subject noun phrase; ! = lexico-semantic mismatch, ǀ = cross-splicing point.

[4]Note that in oral French, a subject with an overt NP ‘The girl’ followed by a pronoun ‘she’ is grammatical (some say the pronoun is obligatory) contrary to written French.

Number mismatches between the depicted subject and the one presented in the auditory stimulus (e.g., the sound file describes ‘she swims’ and the image depicts ‘they swim’) were realized at different sentence positions using cross-splicing techniques (see Tables 2 and 3).

Table 2

Experimental sub-conditions involving liaison (LIAIS) verbs and corresponding visual stimuli.

Visual Stimulus

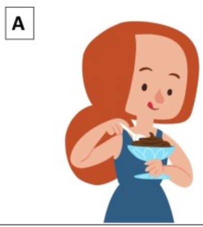


Image A: sample visual stimulus for match conditions (1a-b) and mismatch conditions (2c-d) in the singular. Image B: sample visual stimulus for match (2a-b) and mismatch conditions (1c-d).

Condition	**Number**	**Context**	**Sample auditory stimuli**	
Congruent Morphosyntax	Singular	Neutral	(1a)	*Au dessert ǀ elle aime la mousse au chocolat* ‘For desert ǀ she likes chocolate mousse’
		Subject NP	(1b)	*La fille ǀ elle aime la mousse au chocolat* ‘The girl ǀ she likes chocolate mousse’
	Plural	Neutral	(2a)	*Au dessert ǀ elles‿aiment la mousse au chocolat* ‘For desert ǀ they like chocolate mousse’
		Subject NP	(2b)	*Les filles ǀ elles‿aiment la mousse au chocolat* ‘The girls ǀ she like chocolate mousse’
Incongruent Morphosyntax	Singular	Neutral	(1c)	*Au dessert ǀ elle *aime la mousse au chocolat* ‘For desert ǀ she *likes chocolate mousse’
		Subject NP	(1d)	**La fille ǀ elle *aime la mousse au chocolat* ‘*The girl ǀ she *likes chocolate mousse’
	Plural	Neutral	(2c)	*Au dessert ǀ elles‿*aiment la mousse au chocolat* ‘For desert ǀ they *like chocolate mousse’
		Subject NP	(2d)	**Les filles ǀ elles‿*aiment la mousse au chocolat* ‘*The girls ǀ she *like chocolate mousse’

Note. Critical words are underlined. * = number mismatch, ǀ = cross-splicing point.

Table 3

Experimental sub-conditions involving consonant-final (CONS) verbs, and corresponding visual stimuli.

Visual Stimulus

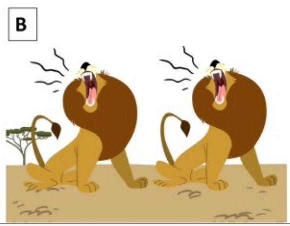


Image A: sample visual stimulus for match conditions (1a-b) and mismatch conditions (2c-d).
Image B: sample visual stimulus for match (2a-b) and mismatch conditions (1c-d).

Condition	Number	Context	Sample auditory stimuli	
Congruent Morphosyntax	Singular	Neutral	(1a)	*En soirée ǀ il rugit dans la savane* In the evening ǀ he roars in the jungle
		Subject NP	(1b)	*Le lion ǀ il rugit dans la savane* The lion ǀ he roars in the jungle
	Plural	Neutral	(2a)	*En soirée ǀ ils rugissent dans la savane* In the evening ǀ they roar in the jungle
		Subject NP	(2a)	*Les lions ǀ il rugissent dans la savane* The lions ǀ they roar in the jungle
Incongruent Morphosyntax	Singular	Neutral	(1c)	*En soirée ǀ il *rugit dans la savane* During evening ǀ he *roars in the jungle
		Subject NP	(1d)	**Le lion ǀ il *rugit dans la savane* The lion ǀ he *roars in the jungle
	Plural	Neutral	(2c)	*En soirée ǀ ils *rugissent dans la savane* In the evening ǀ they *roar in the jungle
		Subject NP	(2d)	**Les lions ǀ il *rugissent dans la savane*

Note. Critical words are underlined. * = number mismatch, ǀ = cross-splicing point.

Two verb types were used; 60 liaison (LIAIS) verbs and 60 consonant-final[5] (CONS) verbs. LIAIS verbs had vowel onsets and were regular 1st conjugation verbs, such as *aimer* 'to-love', which provide no audible cues or disambiguation between 3rd person singular (*aime* [ɛm]) and plural forms (*aiment* [ɛm]). This allowed us to ensure that the only cue for number disambiguation was located at the junction (liaison) between the subject pronoun and the verb,

[5] Within the 60 consonant-final verbs, 24 were subregular verbs and 36 were irregular verbs.

indexed by the presence or absence of the pronoun's plural marker *'s'* [z] (e.g., *elle aime* [ɛlɛm] 'she loves' vs. *elles‿aiment* [ɛl**z**ɛm] 'they love'). Unlike LIAIS verbs, CONS verbs were from the 2^{nd} and 3^{rd} conjugation classes, such as *rugir* 'to-roar', where number distinctions between singular and plural forms are audible on verb endings (e.g., *il rugit* [ilʁyʒ**i**] 'he roars' vs. *ils rugi**ss**ent* [ilʁyʒɪ**s**] 'they roar'). This was the only number cue provided by CONS verbs. A total of 120 verbs (60 LIAIS and 60 CONS) were produced in singular and plural sentences, with both NP and neutral contexts. This resulted in 480 audio files and 960 stimuli: 480 in the congruent condition, and 480 in the incongruent one, where there was a mismatch between the spoken sentence and the picture's verb number.

The 1200 different sentence-picture combinations (240 for conceptual semantics and 960 for agreement) were evenly distributed across four lists (with no sentence repetition within a given list). 300 stimuli sentences with accompanying images were presented to participants in each list (60 for conceptual semantics and 240 for morphosyntax) and were pseudo-randomized (see supplementary material for details). Item versions for each condition were distributed across lists as follows: For semantics, one version of a given verb was included in each list, such that a participant heard one audio file and saw one image (either congruent or incongruent) for each verb. For each LIAIS and CONS verb type two sentence versions of a given verb were included in each list. These sentences were maximally distinct such that they differed in: (1) number (singular vs. plural), (2) context type (neutral vs. subject NP), and (3) congruency (match vs. mismatch with the image), and were presented in different halves of the experiment. This entailed that each subject be presented the same image twice (one match and one mismatch context), but with two completely different audio files.

4.2.3. Procedure

Experimental sessions took place in a quiet room at the UdeM in the third author's lab. Upon arrival, participants read and signed the consent form, after which they completed the Edinburgh Handedness Inventory (Oldfield, 1971) and a language background questionnaire. They were then fitted with an EEG cap, and completed three sub-experiments, all of which used an auditory-visual sentence-picture matching paradigm. The first and second study examined gender-agreement processing (Royle et al, 2013) and word order in French noun phrases. Data from the third experiment are reported here. Total session duration was approximately 3.5 hours, including consent form and other questionnaire completion, working memory test administration, preparation, and clean up.

Participants were seated at a desk at a distance of ~ 40 cm from a computer monitor. Sentences and images were presented using an "Alien learning paradigm", where an alien visited Quebec and was learning French. A story containing filler sentences, images and animations was created, and interspersed throughout the experiment to maintain interest and attention. Participants listened to spoken sentences presented binaurally via insert earphones (ER-1 Insert Earphones, Etymotic Research), while images were presented on the computer monitor. A pause was programmed after every three experimental blocks (60 items).

Participants were instructed to listen to each sentence, while attending to all aspects of grammar and meaning, and judge sentence acceptability in relation to the simultaneously presented image, by pressing one of two keys on a response pad ('acceptable' or 'not acceptable'). In order to avoid laterality effects, the 'acceptable' button was programmed on the right side of the pad for half the participants, and the left side for the other half. Participants were instructed to minimize movement and to keep their eyes open during stimuli presentation. Six practice trials were presented at experiment onset and were excluded from subsequent analyses. At least one researcher or assistant was present throughout the session. EEG recording was monitored throughout, and participants were given feedback about eye blinks and other body movements whenever necessary, in order to reduce artefacts.

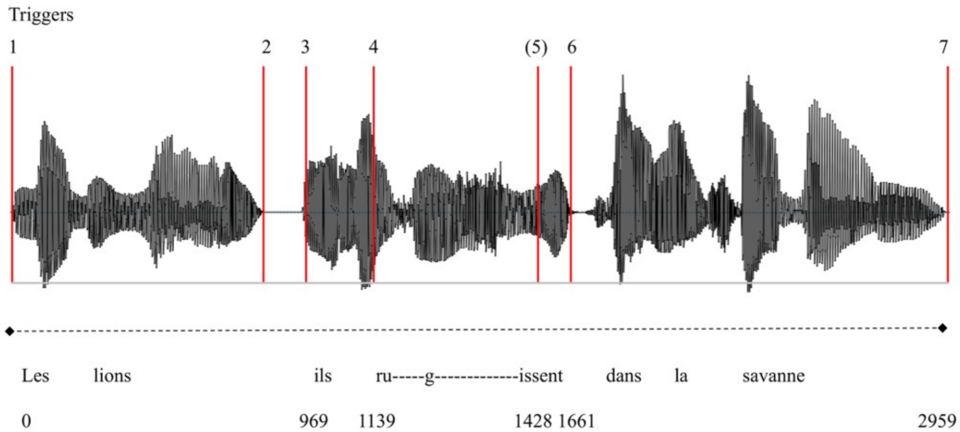
Each trial began with a fixation cross centered on the screen 1000 ms before stimulus presentation. The image was presented 500 ms before sentence onset, and stayed on screen until the auditory stimulus ended. After the sentence, a blank screen appeared for 1000ms, then a response prompt ('???') appeared on the screen until a response button was pressed. This was followed by a fixation cross for 1000 ms, during which subjects were instructed to blink their eyes before the next trial began, and a blank screen for 1000 ms

4.2.4. Analysis time-locking

In order to quantify the time course of number mismatch and lexical-semantic effects, our analyses were time-locked to relevant lexical-semantic and morphophonological cues (Steinhauer and Drury, 2012), using triggers at relevant speech signal positions. Figure 1 depicts an example waveform for the sentence *Le lion, il rugit dans la savane* 'The lion, he roars in the jungle' as well as its trigger points. Analyses presented in this paper use triggers 1 (sentence onset) and 4 (verb onset).

Figure 1

Example waveform (in ms) of an auditory stimulus



Note. Example waveform (in ms) of an auditory stimulus for the sentence *Les lions, ils rugissent dans la savane*. The red lines represent the various cue points, called 'triggers', measured in the audio file. Trigger 1 = sentence onset; Trigger 2 = context phrase offset; Trigger 3 = pronoun clitic onset; Trigger 4 = verb onset; Trigger 5 = onset of verb-final consonant (only Type 2 verbs); Trigger 6 = verb offset; Trigger 7 = sentence offset.

4.2.5. EEG recording and data analysis

The EEG was recorded continuously with a 500 Hz sampling rate from 64 cap-mounted electrodes (WaveGuard caps, ANT; Enschede, NL) placed according to the extended International 10/20 System. The electrodes used for recording covered frontal, central, parietal, temporal and occipital lobes (FP1, FP2, F3, F4, F7, F8, Fz, C3, C4, Cz, P3, P4, Pz, T3, T4, T5, T6, O1, O2, Oz). All impedances were maintained below 5 kΩ and were checked every 45 minutes throughout the experiment. The EEG was amplified using an ANT Neuro Ego™ sports amplifier referenced to the CPz electrode. All subsequent EEG/ERP data processing steps and analyses were carried out using EEProbe software package (ANT; Enschede, The Netherlands) and statistical analyses were performed in R (RStudio Team, 2015, Integrated Development for

R. RStudio, Inc., Boston, MA[6]) using the Easy analysis and factorial experiments visualization package (Lawrence, MA. 2011, R package version 4.4-0.[7])

Offline, raw data were re-referenced to linked mastoids and filtered using a Gaussian bandpass filter of 0.3 to 40 Hz. Trials contaminated with eye blinks or other artefacts were rejected using a 30 µV criterion. All uncontaminated trials were entered into the final analysis. Using a 600 ms pre-stimulus baseline interval, single-subject EEG waveforms per condition were averaged separately over 2100 or 3100 ms epochs (-600 to 1500 or 2500 ms), time-locked to the relevant critical word onset (underlined words in Tables 1-3 above) and entered into grand average ERPs. After artifact rejection, an average of 48/60 trials for semantic mismatches and 192/240 trials for number mismatches were analyzed per participant. Based on visual inspection and the previous literature, we identified representative time-windows for statistical analyses of leixical-semantic and number mismatches, during which ERP components were quantified as the mean EEG signal voltage (in µV).

In all analyses, we compare mismatch conditions to their corresponding match conditions presenting the exact same spoken sentence but with a different picture. For example, a number mismatch analysis for singular sentences compares singular spoken sentences with subject NPs, combined with a corresponding picture showing one agent (match condition) or with a similar picture showing two agents (mismatch condition). ERP analyses for midline electrodes and lateral electrodes were performed separately. At midline electrodes, global ANOVAs for the semantic condition included 2 factors: CONDITION (2 levels: mismatch vs. match), and ELECTRODE position (4 levels: Fz, Cz, Pz, and Oz). At lateral electrodes, the global ANOVA

---

[6] http://www.rstudio.com/
[7] http://CRAN.R-project.org/package=ez

included four factors: CONDITION (2 levels: mismatch vs. match), HEMISPHERE (2 levels: right vs. left), ANTERIORITY (3 levels: anterior vs. central vs. posterior), and LATERALITY (2 levels: lateral vs. medial). For the *number* mismatch conditions, two additional factors were included for both analyses: CONTEXT (neutral vs. subject NP) and NUMBER (singular vs. plural). Greenhouse-Geisser corrections were applied in order to address potential violations of sphericity. In these cases, the original degrees of freedom and corrected probability levels are reported. A hierarchically-organized analysis of variance was pursued whereby only theoretically relevant interactions (i.e., CONDITION effects and their interactions with scalp distribution effects) and attendant post-hoc analysis results are reported. Given that the ERP effects of interest are generally observed close to the midline rather than at more lateral recording sites, 12 representative electrodes are used to illustrate effects, while head maps for difference waves cover the whole scalp.

Arcsine transformed accuracy data from acceptability judgments were analyzed using repeated-measure ANOVAs, computed separately for semantic and number conditions. The global ANOVA for number mismatches included four factors with 2 levels each: CONDITION, CONTEXT, GENDER, and NUMBER.

4.3. Results and interim discussions

Following a reviewer's suggestion, we first present behavioral data (section 4.3.1), followed by ERP results and discussion for lexico-semantic mismatches (section 4.3.2), and finally results and discussion for number mismatches (section 4.3.3).

4.3.1. Behavioral data results

Accuracy for acceptability judgments for *lexical-semantic* conditions were nearly at ceiling for both match and mismatch sentences (see Table 4), and a global ANOVA indicated no

CONDITION effect ($p < 1$). Global ANOVAs for *number* mismatches on LIAIS verbs revealed significant main effects of CONDITION ($F(1,21) = 6.39$, $p = .0196$) in favor of matches, and NUMBER ($F(1,21) = 5.67$, $p = .0269$) in favor of the plural (Singular: Mean 93.6, SD = .045; Plural: Mean = 95.7, SD = .048), qualified by interactions for CONDITION×NUMBER ($F(1,21) = 8.97$, $p = .0069$), CONDITION×CONTEXT ($F(1,21) = 5.90$, $p = .0242$, and NUMBER×CONTEXT ($F(1,21) = 9.60$, $p = .0054$). All these interactions are primarily driven by lower rejection rates for singular mismatches in neutral contexts (in **bold**, Table 4), where number disambiguation was realized by the lack of a plural marker at the liaison. See section 3.3.2 for further discussion. A global ANOVA for CONS verbs revealed that these differed significantly by CONDITION ($F(1,21) = 4.52$, $p = .0455$), but no other significant effects were found. Mismatches were responded to less accurately than matches.

Table 4

Accuracy means (and standard deviations) for audio-visually matching and mismatching trials.

Conditions	Match	Mismatch
Lexico-semantics	93.9 (.060)	92.5 (.070)
Morphosyntax. Number: Liaison verbs	96.3 (.035)	93.0 (.066)
SINGULAR: NP CONTEXT	96.1 (.069)	94.2 (.082)
SINGULAR: NEUTRAL CONTEXT	97.6 (.036)	**86.5 (.105)**
PLURAL: NP CONTEXT	94.4 (.074)	94.7 (.079)
PLURAL: NEUTRAL CONTEXT	97.2 (.046)	96.8 (.087)
Morphosyntax. Number: Consonant-final verbs	94.8 (.039)	91.6 (.077)

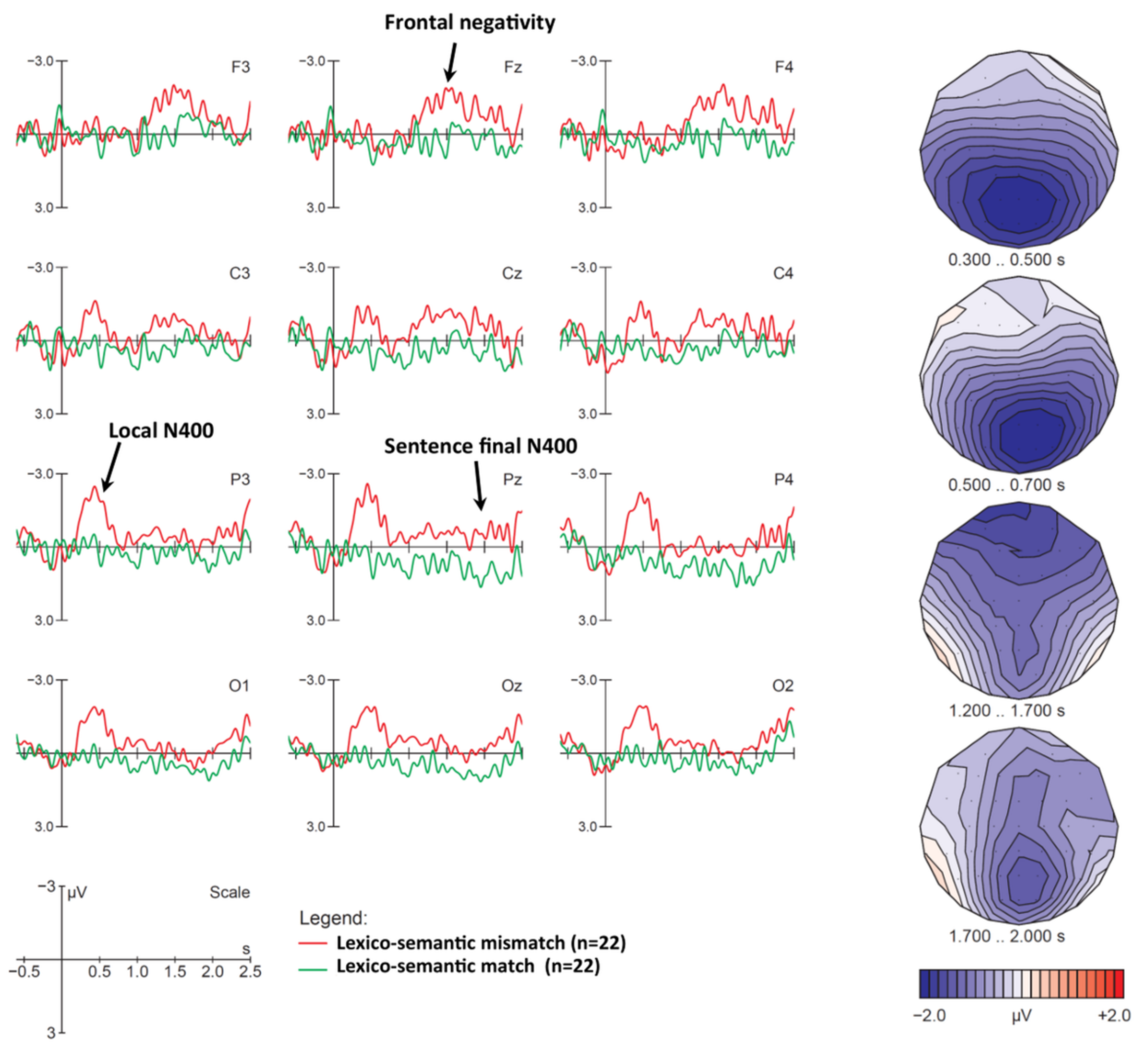
Note. Sub-conditions (for number and context) are listed only where statistical analyses indicated different patterns (i.e., for LIAS verbs).

4.3.2. ERP for lexico-semantic mismatches

As depicted in Figure 2, compared with the match condition, the semantic mismatch condition elicited a series of negativities across both context conditions at verb onset. First, we observe a posterior N400-like negativity between roughly 300–700 ms. Secondly a subsequent negative deflection emerges around 1200 ms and lasts until 2000 ms, it shows a frontal distribution until 1700 ms and becomes more posterior afterward. Recall that the verb was always followed by an object noun phrase (NP) or a prepositional phrase (PP) that ended the sentence, and that nouns within these phrases also mismatched with the depicted information (see Table 1 for an example). On average, verbs ended 550 ms after onset, and participants heard the NP/PP between 600 and 1800 ms. Based on this time course, we analyzed the negativities in five different time windows: 300–500 ms for the core N400, 500–700 ms for the extended N400, 700–1100 ms for the interval that did not elicit effects, 1200–1700 ms for the negativity related to the NP/PP mismatch, and 1700–2000 ms for a presumed sentence-final N400-like negativity. Statistical analyses for all time windows, separately for lateral and midline electrodes, are summarized in Table 5.

Figure 2

ERP effects for the lexico-semantic mismatches.



Note. Displayed are grand-average ERPs at midline and eight lateral electrodes, as well as voltage maps illustrating the difference waves, for all participants, time-locked to the onset of the critical verb using a baseline of -600 to 0 ms. The vertical bar marks the onset of the critical verb. On average the verb ended 550 ms after onset; between 600 and 1800 ms participants heard a NP/PP, which included a second semantic mismatch and ended the sentence. Compared with the correct match condition (green line), the semantic mismatches (red line) elicited a large extended N400 between 300 and 700 ms, followed by a frontal negativity during the NP or PP (1200–1700 ms), and a subsequent posterior sentence wrap-up N400 between 1700–2000 ms. Negative polarity is plotted upwards. Voltage maps represent difference waves (violation minus control), with negativities in blue and positivities in red. For illustration purposes only, ERP plots have been 10 Hz low-pass filtered.

Table 5

Global repeated measures ANOVAs for lexico-semantic conditions at time-windows of interest.

		(N400)			Late negativity	Wrap up effects
	df	300–500	500–700	700–1100	1200-1700	1700-2000
LATERAL ELECTRODES						
CONDITION	(1, 21)	—	—	—	7.14**	—
CONDITION × ANTERIORITY	(2, 42)	5.26***	8.63***	—	6.28**	—
FRONTAL: CONDITION	(1, 21)	—	—	—	9.82**	—
CENTRAL: CONDITION	(1, 21)	—	—	—	5.02*	—
POSTERIOR: CONDITION	(1, 21)	8.08**	8.08**	—	—	—
CONDITION × LATERARLITY	(1, 21)	9.59**	5.34*	—	4.77*	—
MEDIAL: CONDITION	(1, 21)	5.44*	3.60†	—	7.06*	—
LATERAL: CONDITION	(1, 21)	—	—	—	5.26*	—
CONDITION × ANT × CONTEXT	(2, 42)	—	5.26*	—	—	—
NP CONTEXT: CON × ANT	(2, 42)	—	13.16***	—	—	—
NP CONTEXT ANT: CON	(2, 42)	—	8.09**	—	—	—
CONDITION × LAT × ANT	(2, 42)	—	—	—	4.51*	—
CENTRAL: CON × LAT	(1, 21)	—	—	—	4.92*	—
CENTRAL, MEDIAL: CON	(1, 21)	—	—	—	5.67*	—
POSTERIOR: CON × LAT	(1, 21)	—	—	—	10.13**	—
LATERAL: CON × ANT	(2, 42)	—	—	—	11.38***	—
LATERAL, FRONTAL: CON	(1, 21)	—	—	—	11.21***	—
CONDITION × LAT × CONT	(2, 42)	—	—	—	5.69*	—
NEUTRAL: CONDITION × LAT	(1, 21)	—	—	—	7.80**	—
NEUTRAL, MEDIAL: CON	(1, 21)	—	—	—	8.55**	—
MIDLINE ELECTRODES						
CONDITION	(1, 21)	5.56*	—	—	10.26***	7.35**
CONDITION × ELECTRODE	(3, 63)	6.34*	10.79***	—	—	—
PZ: CONDITION	(1, 21)	9.29***	7.35**	—	—	—
OZ: CONDITION	(1, 21)	9.21***	12.85***	—	—	—

Note. Analyses at trigger 4. Only significant results and trends are presented. Con = Condition, Ant = Anteriority, Lat = Laterality, Cont = Context; † $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Significant interactions in the global ANOVA were decomposed to identify scalp electrodes displaying the strongest condition differences. In both the 300–500 ms and 500–700 ms time windows, the most dominant and consistent effects included CONDITION× ANTERIORITY interactions at both lateral and midline electrodes, as well as a CONDITION× LATERALITY interaction at lateral electrodes. Decomposing these interactions confirmed that the N400 reached significance only at posterior electrodes at or near the midline (Pz and Oz, and posterior medial electrodes). As expected, for the 700–1100 ms time-window, we found no significant main effects or interactions involving CONDITION. As can be seen in Figure 2 (e.g., at Pz), the absence of an effect in this contrast cannot be attributed to the presence of a P600 that may have cancelled out any ongoing negativities due to component overlap. In fact, there is not the slightest indication of a positive dip that could point to a “hidden” P600, including at posterior electrodes where P600s are usually found.

A global ANOVA for time-window 1200–1700 ms yielded a significant CONDITION effect at midline and lateral electrodes, as well as CONDITION×ANTERIORITY, CONDITION× LATERALITY, CONDITION× LATERALITY×ANTERIORITY, and CONDITION×LATERALITY×CONTEXT interactions. The first three interactions indicate that this broadly distributed late negativity is most prominent at frontal electrodes and along the entire midline, whereas it gradually decreases at more lateral and posterior sites over both hemispheres (see voltage map). Finally, decomposing the interaction involving CONTEXT, we found that the negativity was more broadly distributed in the NP context, but limited to medial electrodes in the neutral one. Global ANOVAs for the sentence “wrap-up” effect in the 1700–2000 ms time-window yielded a CONDITION main effect in the midline with no other interactions.

4.3.2.1. Discussion for N400 effects

Lexico-semantic mismatches on verbs were reliably detected by participants and elicited a large N400 component, as expected. Importantly, our study focused on mismatches involving verbs/actions, and not nouns/objects as in Royle et al (2013) and other previous studies. We have therefore demonstrated that an N400 can be reliably elicited in adult French native speakers in response to verb-action mismatches. We believe that these require more complex cognitive matching processes than noun-object pairings, as they involve syntactic and thematic relations between a verb and its arguments. For example, in order to appropriately illustrate the ditransitive verb *give*, one must include an agent, a patient, and a beneficiary.

After the classic N400 time-window (300–500 ms), the N400 continued until 700 ms post verb-onset. There are various possible interpretations for this finding. First, mismatches involving verbs rather than nouns may require more complex processing. Secondly, in auditory studies, the N400 sometimes shows a longer duration due to word variability across trials (Holcomb and Neville, 1990). Thirdly, extended N400s with durations up to 700 ms have been discussed as reflections of additional post-lexical integration. The relevant discussion concerns the N400's functional interpretation, and whether it simply reflects automatic expectancy-based processing (i.e., lexical access typically between 300–500 ms, Federmeier, 2007; Kutas et al., 2006; Lau et al., 2008) or whether it also reflects controlled post-lexical integration (i.e., spoken word integration into a higher-order meaning representation after 500 ms, e.g., Brown and Hagoort, 1993; Holcomb, 1993; Steinhauer et al., 2017). Fourthly, 2/3 of our verbs were immediately followed by a direct object, which, in this condition, also mismatched with the visual stimulus, and may therefore have elicited a second N400. Note that the negativity's scalp distribution between 500–700 ms resembled the N400 preceding it, such that it is impossible to rule out any of these explanations without additional analyses beyond the scope of this paper.

4.3.2.2. Discussion for sustained frontal and posterior negativities

Following N400 effects, we observed late sustained negativities, the first between 1200 and 1700 ms with a frontal distribution, and the second between 1700 and 2000 ms with a broad distribution, but a central-parietal maximum consistent with an N400. The frontal negativity was elicited while direct objects (NP) or prepositional phrases (PP) were being processed. Both the NP and the noun in the PP also mismatched with the picture (i.e., one sees a woman singing on a stage but hears *'she swims in the public pool'*, see Table 1). A comparison of this condition and the number mismatch conditions, where no incongruencies were present between the NP/PP in erroneous and correct sentences (see Figures 5 and 6 below), shows that we observe a sustained negativity between 1200 and 1700 ms only in the lexico-semantic mismatch condition, suggesting that it is related to this additional semantic mismatch. However, its frontal distribution is not typical of an N400 and may point to a combination of mismatch effects proper and frontal expectancy effects reflecting anticipation of an additional semantic mismatch. Similar effects have been found for anticipation of a predictable comma likely to render a sentence ungrammatical, and was interpreted as a contingent negative variation (CNV, Steinhauer, 2003). We interpret the late portion of the negativity as a potential “sentence wrap-up effect”, which we discuss in the section Sentence-final negativities and wrap-up effects (4.3.4.8).

4.3.2.3. Discussion for P600 effects

Recall that the P600 has sometimes been elicited by semantic anomalies in conjunction with the N400, notably in an cross-modal mismatch paradigm (Royle et al., 2013), but also in purely auditory ones (Hagoort, 2003), and in reading studies (Steinhauer et al., 2010), and has therefore been argued to reflect mental monitoring and processing load related to language reanalysis (i.e., it is not specific to grammatical processing; Kolk et al., 2003; Steinhauer and Connolly, 2008; van de Meerendonk et al., 2009). Others have argued that these positivities are tightly linked to acceptability judgment tasks, potentially as a linguistic variant of the P300 component (Coulson et al., 1998; Sassenhagen et al., 2014; Friederici et al., 2001). The absence of positivities in the lexico-semantic condition, despite our use of a judgment task, may be explained by our particular mismatches. First, as reflected by the subsequent frontal negativities, participants seemed quite engaged in anticipating and processing additional semantic mismatches in the following NPs and PPs, and may not have categorized the sentence as unacceptable when encountering semantic mismatches on verbs. Another possibility is that semantic mismatches realized on verbs do in fact involve more complex conceptual-semantic processing than those realized on nouns and may draw attention away from whatever processes may elicit positivities found on nouns. As we are not aware of any other ERP studies using verb/action mismatches, this would need to be further investigated. Finally, P600s are certainly not a consistent finding for conceptual mismatches; the motivation for explaining their absence is primarily based on their presence in a recent study from our lab that used a very similar cross-modal paradigm (Royle et al., 2013). Perhaps the most important point is that the absence of a P600 in our semantic mismatch condition contrasts with the P600s observed in other mismatch conditions that we will discuss next.

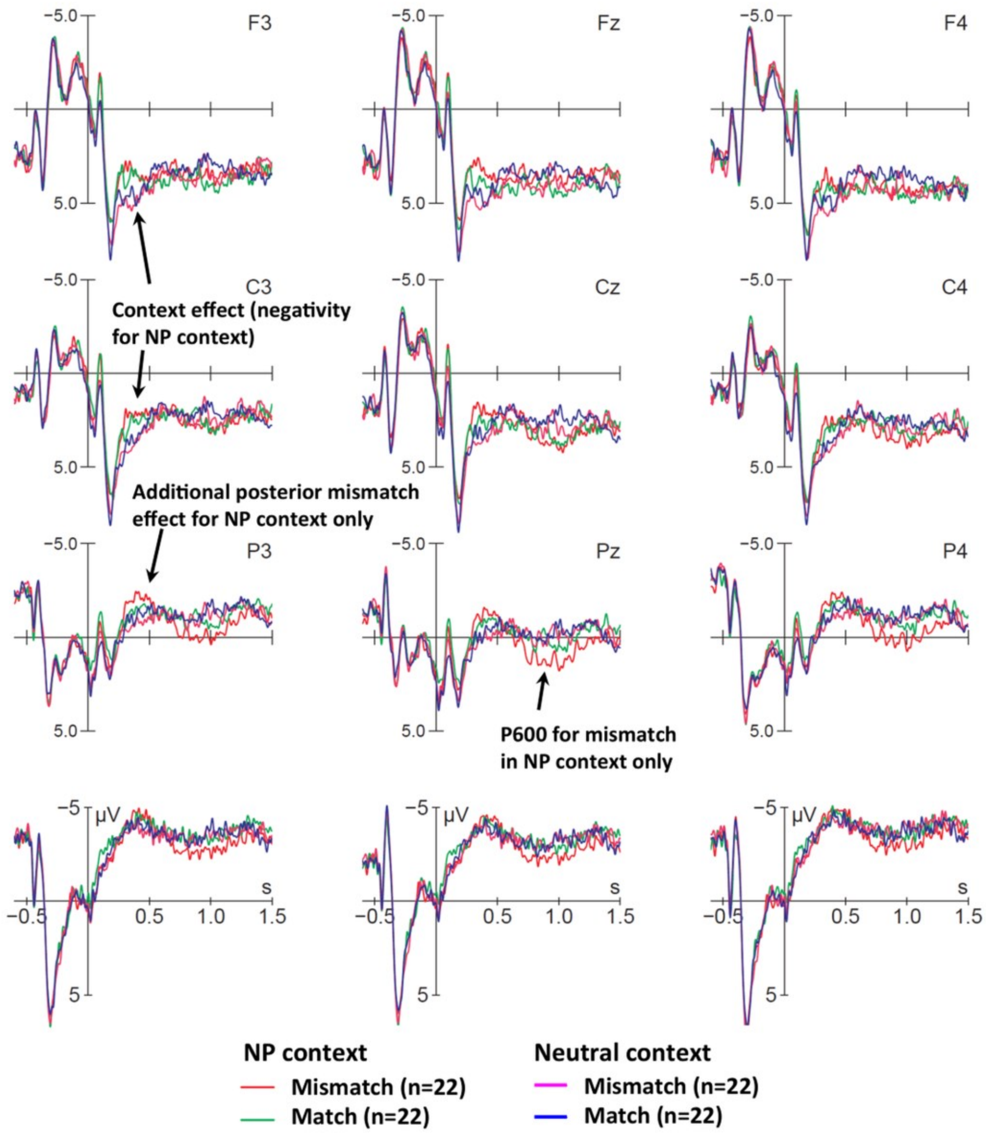
4.3.3. ERPs for number mismatches at sentence onset

At sentence onset we observed distinct ERP patterns for neutral contexts (with no disambiguation at this point) and NP contexts, where the NP either matched or not with the picture in number at the determiner (*le/la/les* 'the.M.SG/F.SG/PL'). The distinction between LIAIS and CONS verbs does not play a role at this point, such that we can collapse across these conditions, which we did. Figure 3 displays match and mismatch conditions for both NP and neutral contexts, collapsed across singular and plural sub-conditions. Recall that the mismatch in neutral contexts happens only downstream on the verb and is, therefore, not yet expected to elicit mismatch components. The first 900 ms (-600 to 300 ms) are largely dominated by visual onset components (most prominently at occipital electrodes) for pictures (presented at -500 ms) and by auditory onset components (most prominently at fronto-central electrodes) for spoken sentences (starting at 0 ms), respectively. As can be seen, all conditions are virtually indistinguishable up to 300 ms after sentence onset, at which point the first context-effect emerges.

NP contexts, compared to neutral ones, elicited an early slightly left-lateralized fronto-central negativity (300–450 ms) after determiner onset. In the same time-window, we observe an additional enhanced negativity for NP context mismatches, which is followed by a P600 (700–1200 ms). We will show how singular and plural mismatches in NP contexts contribute to this pattern. In neutral context conditions – as expected – no clear differences are visible, as confirmed by the absence of significant effects in all time-windows discussed below (see also Table 6). We return to neutral contexts at later sentence positions – at verb onset – where they are disambiguated.

Figure 3

Early ERP effects of context and number mismatches at sentence onset.



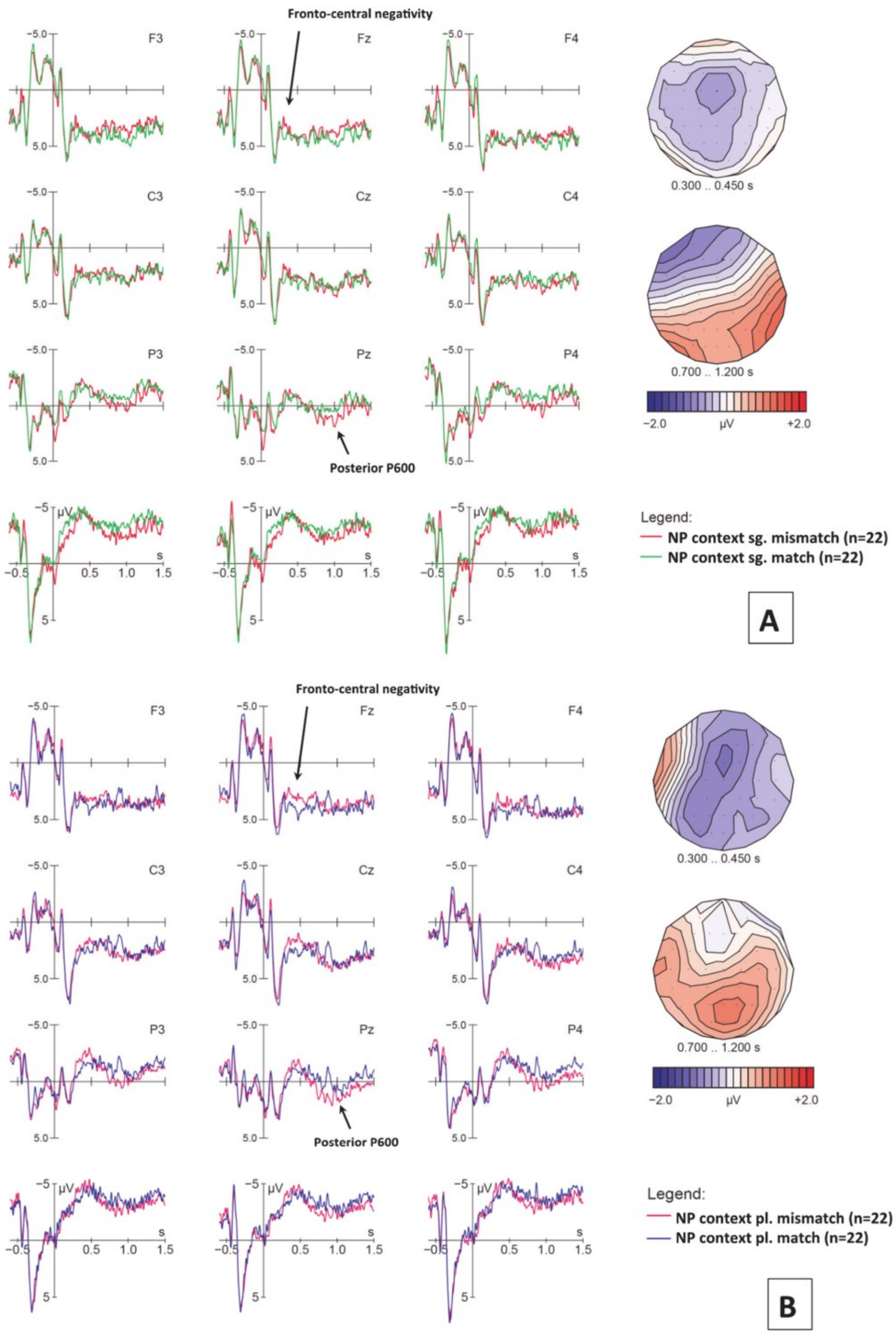
Note. Displayed are grand-average ERPs at midline and lateral electrodes for all participants, time-locked to the onset of the determiner (vertical bar) with a baseline of -600 to 0 ms. Compared with neutral context correct (blue), and neutral context mismatch (magenta), the NP context correct condition (green) and the NP context mismatch condition (red) elicited an early negativity (300–500 ms). Furthermore, number mismatches with NP context display a small increased negativity (between 300-450 ms) and a large positivity between 700–1200 ms. The two neutral conditions will be disambiguated further downstream at the verb and do not yet show differences at sentence onset.

4.3.3.1. ERPs for singular and plural mismatches in NP contexts

For sentences with *singular* NPs, we observe a small fronto-central negativity in the N400 time-window, followed by a posterior P600 in the mismatch condition between 700 and 1200 ms after sentence onset (see Figure 4A). In the *plural* contrast (Figure 4B), we see a similar biphasic pattern for mismatches, however, the fronto-central negativity appears slightly larger and seems to extend more clearly to left posterior electrodes. Statistical analyses for sentence-initial positions are summarized in Table 6.

Figure 4

Early effects of number mismatches in NP contexts, for (A) singular and (B) plural NPs at sentence onset.



Note. ERPs are time-locked to the onset of the determiner (vertical bar) with a baseline of -600 to 0 ms; voltage maps illustrate the difference waves of relevant effects.
(A) Singular mismatches (red) elicited a small fronto-central negativity in the N400 time-window relative to singular matches (green), as well as a parietal P600. **(B)** Plural mismatches (magenta) elicited a larger N400 as well as a parietal P600, as compared to plural matches (blue). Voltage maps of these effects (mismatch minus control) show that singular and plural mismatches elicited quite similar components.

Table 6

Global repeated measures ANOVAs for sentence onset effects at time-windows of interest.

		(N400)	(P600)
	df	300–450	700–1200
LATERAL ELECTRODES			
CONDITION	(1, 21)	—	—
CONTEXT	(1, 21)	29.03***	2.99†
CONDITION × CONTEXT	(1, 21)	5.32*	—
CONDITION × LAT × CONT	(1, 21)	6.70*	—
CONDITION × ANT × CONT	(2, 42)	—	7.95**
NP CONTEXT: ANT × COND	(2, 42)	—	9.56***
NP CONTEXT, POST: COND	(1, 21)	—	11.69***
NP CONTEXT ONLY			
CONDITION × ANTERIORITY	(2, 42)	—	7.89**
POSTERIOR: CONDITION	(1, 21)	—	10.95*
CONDITION × ANT × HEM × NUM	(2, 42)	—	4.56*
POSTERIOR: CONDITION × HEM × NUM	(1, 21)	—	10.95*
CON × ANT × HEM × NUM × LAT	(2, 42)	—	5.81*
LEFT HEM: CONDITION × ANT	(2, 42)	—	7.80**
LEFT HEM: CONDITION × ANT × NUM	(2, 42)	—	5.97*
LEFT HEM: SG: CONDITION × ANT	(2, 42)		10.49***
LEFT HEM: SG: FRONT: CONDITION	(1, 21)	—	3.23†
LEFT HEM: SG: POST: CONDITION	(1, 21)	—	5.03*
LEFT HEM: PL: CONDITION × ANT	(2, 42)		3.90*
LEFT HEM: CONDITION × ANT × NUM × LAT	(2, 42)	—	4.25*
MIDLINE ELECTRODES			
CONDITION	(1, 21)	—	—
CONTEXT	(1, 21)	20.56***	9.58**
CONDITION × CONTEXT	(1, 21)	9.78**	—
NP CONTEXT: CONDITION	(1, 21)	4.43*	—
CONDITION × ELEC × CONTEXT	(3, 63)	—	8.52***
PZ: CONDITION	(1, 21)	—	8.44**
PZ: CONDITION × CONTEXT	(1, 21)	—	6.22*
PZ: NP: CONDITION	(1, 21)	—	12.10***
OZ: CONDITION × CONTEXT	(1, 21)	—	10.92*
OZ: NP: CONDITION	(1, 21)	—	9.34**
NP CONTEXT ONLY			
CONDITION	(1, 21)	4.43*	9.14***
PZ: CONDITION	(1, 21)	—	11.35***
OZ: CONDITION	(1, 21)	—	8.36**

Note. Analyses at trigger 1. Only significant results and trends are presented. Ant = Anteriority, Con = Condition, Cont = Context, Elec = Electrode, Front = Frontal, Hem = Hemisphere, Lat = Laterality, Num = Number, Pl = Plural, Post = Posterior, Sg = Singular, † $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Global ANOVAs in the 300–450 ms time-window yielded a highly significant CONTEXT main effect. Mismatch effects were reflected by CONDITION×CONTEXT interactions in midline and lateral electrodes, as well as a CONDITION×CONTEXT×LATERALITY interaction in lateral electrodes. These interactions confirmed that the negativity for visuo-auditory number mismatches was limited to disambiguating NP contexts, and was largely limited to medial electrodes. Surprisingly, the absence of significant ANTERIORITY and NUMBER interactions suggested that (a) the apparent frontal focus of the negativity was not reliable across subjects and (b) the apparent differences in size and scalp distribution of negativities between singular and plural conditions (Figures 4A vs 4B) were not meaningful. Statistically, there was only a broadly distributed negativity in both singular and plural mismatches with NP contexts.

In the P600 time window (700–1200 ms), global ANOVAs yielded a significant CONDITION× ELECTRODE×CONTEXT interaction at midline electrodes, and CONDITION×CONTEXT and CONDITION× LATERALITY×CONTEXT interactions in lateral electrodes (see Table 6). Decomposing these interactions confirmed that the P600 had a posterior distribution and was limited to number mismatches in NP contexts. While this P600 was consistent across singular and plural at midline electrodes (significant CONDITION main effect at Pz and Oz), additional interactions with factor NUMBER and topographical factors at lateral electrodes indicated that only for singular mismatches the P600 time-window also showed a (non-significant) frontal negativity over the left hemisphere. Overall, both singular and plural mismatches with NP contexts elicited consistent P600s that lasted until 1200 ms.

Note that this relatively long P600 duration means that this effect was still present when the verb was presented (average verb onset at 1140 ms, SD = 149 ms) and would have contaminated baselines and ERP analyses time-locked to verb onset (cf., Steinhauer & Drury, 2012). For these reasons, we refrained from analyzing the NP-context conditions at the verb, even though it would have been interesting to see whether additional disambiguating information elicited more mismatch effects further downstream.

4.3.3.2. Discussion for sentence-initials effects

Independent of mismatches, context manipulations at sentence onset elicited a larger negativity for NP contexts between 300–450 ms after sentence onset: this was likely triggered by the first word. Both NP contexts and neutral contexts started with function words (e.g., *Au dessert* ‘at-the desert’ in neutral context / *La/les fille/s* ‘The girl/s’ in NP context) for which N400 effects are rather atypical. In addition, the context-driven negativity had a more frontal distribution than a classic N400. We speculate that this context main effect may reflect enhanced alertness once participants had identified that a sentence started with a determiner and could, therefore, provide the first disambiguating task-relevant cue.

Interestingly, determiner *mismatches* elicited an additional, more broadly distributed negativity in virtually the same time-window, which was followed by a posterior P600, for both singular and plural mismatches. The mismatch negativity could be interpreted either as a lexical prediction effect (i.e., an N400, Tanner and Van Hell, 2014; BSS2018) or an effect of reference resolution (i.e., an N-ref component, e.g., Van Berkum et al., 1999). In the first scenario, participants would expect a specific determiner coherent with the number (and gender) of depicted potential subjects, and process a mismatch as a lexical (or phonological) error. In the second scenario, participants might wonder, when there are multiple potential subjects, who *la* *fille* ‘the girl’ refers to. However, reference resolution effects only seem to make sense – and have only been reported – for singular nouns where contexts provide multiple potential referents, while we found no statistical differences between our singular and plural conditions and, moreover, we found them at the determiner rather than the noun. For these reasons we believe that this negativity reflects a mismatch for specific predictions. Our finding is reminiscent of that by DeLong and colleagues (2005) who reported an N400 on determiners for unexpected sentence continuations after a highly constraining context (e.g., <u>*an*</u> *airplane* rather than <u>*a*</u> *kite* after ‘… the boy went outside to fly _’). Whether this effect is primarily lexical or phonological in nature remains unclear.

The following P600-like positivity in our data may either reflect (a) an immediate categorization of the sentence as unacceptable (Sassenhagen et al, 2014) or (b) cross-modal integration of conflicting number information as in previous morphosyntactic (dis-)agreement studies, possibly linked to structural disambiguation or revisions (see e.g., Molinaro et al., 2011a, for a review), or both. In line with our previous work and the literature (e.g., Friederici et al., 2001; Royle et al., 2013; Steinhauer & Connolly, 2008), we maintain the view that the P600 typically reflects multiple cognitive processes and comprises multiple subcomponents. A P600 account involving structural (rather than purely lexical) mismatches or revisions would imply that participants in our study syntactically integrated the determiner with the subsequent noun, which was phonologically compatible with both a singular and a plural form (*fille/s* [fij]). However, a picture of two girls would have suggested (and pre-activated) a plural referent, which then mismatched with the spoken singular determiner (*la* ‘the.SING.FEM’), thereby resulting in a traditional number agreement violation (i.e., *la *filles*). Given that these early-disambiguating contexts were followed by additional information disambiguating subject number on the verb, one might expect higher confidence (and thus higher accuracy) in grammaticality ratings compared to sentences with neutral contexts. However, as discussed above (see also Table 4), this was not the case, supporting immediate categorization at the first available cue. We anticipate that this pattern may be different in children, especially those with language impairment, who are currently being tested with this same paradigm.

For obvious reasons, number mismatch effects at sentence-*initial* words (as in our study) are absent from the previous literature as they can only be created in relation to a previously presented context (here: a picture). Overall, it is remarkable that this sentence-initial number mismatch elicited an N400-P600 pattern previously found for morpho-syntactic agreement violations. It suggests that nonlinguistic visual information from the environment can be immediately used (in less than 500 ms) to make strong predictions about appropriate linguistic representations, or that "feature checking" processes are not constrained to linguistic representations. The elicitation of a P600 at this early position in a sentence is clearly compatible with accounts of "conflict monitoring" (Kolk et al., 2003) and "well-formedness categorization" (Sassenhagen et al., 2014), but more difficult to explain in terms of a structural "reanalysis" (Friederici, 2002).

4.3.4. ERPs for number mismatches on verbs

We will now turn to mismatch effects at target verbs in neutral contexts. At sentence onset, LIAIS and CONS verbs did not differ, but at trigger 4 (verb onset) they did, because for LIAIS verbs, number disambiguation is available at verb onset (e.g., *elles[z]aiment* 'they like'), while for CONS verbs, this information is available only at the verb final phoneme (e.g., *ils rugissent* [ryʒɪ**s**] 'they roar'). We will first focus on LIAIS verbs and then turn to CONS ones and consider only neutral contexts because these are the ones being disambiguated for the first time on the verb.

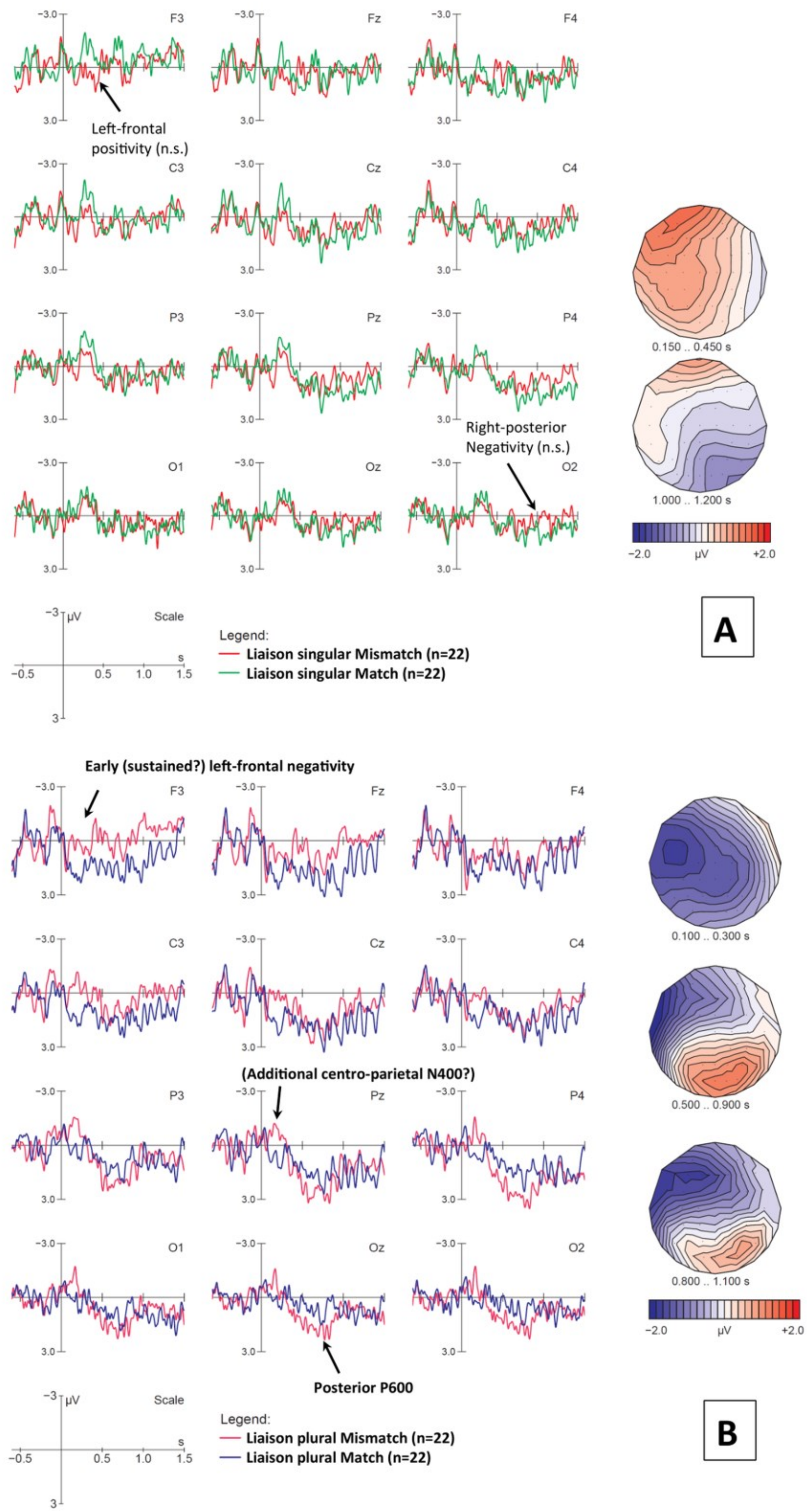
4.3.4.1. ERPs for liaison verbs at verb onset at Trigger 4

As with sentence initial effects, we analyzed singular and plural violations separately. Figure 5A shows number mismatches time-locked to singular LIAIS verbs. In this comparison we did not observe the expected pattern but rather an apparent early left-anterior positivity between 150–450 ms after verb onset, and a posterior right-lateralized late negativity between 1000–1200 ms. However, as seen in Table 7, global ANOVAs on singular LIAIS verbs in neutral conditions yielded no significant effects involving CONDITION at either the midline or lateral electrodes. (Note that the very early left-frontal positivity was partly driven by one participant's enhanced horizontal eye movements in this condition only, resulting in a polarity inversion of this difference between left-anterior and right-anterior electrodes – especially F7 and F8. Analyses excluding this data set did not change results, however. For consistency, we decided to present ERP data including this data set). Overall, our analyses did not point to any consistent ERP pattern for these number mismatches. Recall that this was also the condition with the lowest overall accuracy rate in our mismatch conditions (Table 4).

As illustrated in Figure 5B**,** for plural mismatches we observed an early left-lateralized fronto-central negativity between 100 and 300 ms, followed by a posterior P600-like positivity (500–900 ms), which then seems to be followed by a second late frontal and somewhat left-lateralized negativity from approximately 800–1200 ms. In fact, when inspecting the left-anterior electrode F3 alone, the patterns looks like a sustained early negativity, starting around 100 ms and lasting until approximately 1400 ms. Statistical analyses are presented in Table 7.

Figure 5

ERP effects for number mismatches at liaison verbs (A) for singular and (B) for plural verbs.



Note. Displayed are grand-average ERPs at midline and lateral electrodes for all participants, time-locked to the onset of the liaison using a baseline of -600 to 0 ms. The vertical bar marks the onset of the liaison. **(A)** For singular verbs, neither the early frontal positivity between 150 and 450 ms nor the posterior negativity (1000-1200 ms) reached significance. **(B)** Compared to the correct control condition (blue lines), plural mismatches (magenta lines) show early negativities (100-300 ms), followed by a posterior P600 (500-900 ms). After the end of the P600, a negativity seems to re-emerge at frontal and central electrodes (third voltage map).

Table 7

Global repeated measures ANOVAs for liaison verbs for both singular and plural at time-windows of interest.

		(LAN)	(P600)
	df	100–300	500-900
Singular verbs			
Lateral electrodes			
Condition	(2, 42)	—	—
Condition × Anteriority	(2, 42)	—	—
Midline electrodes			
Condition	(3, 36)	—	—
Condition × Electrode	(3, 36)	—	—
Plural verbs			
Lateral electrodes			
Condition	(1, 21)	6.39*	—
Condition × Laterality	(1, 21)	6.12*	—
Medial: Condition	(1, 21)	7.22**	—
Lateral: Condition	(1, 21)	3.63†	—
Condition × Anteriority	(2, 42)	—	6.66**
Condition × Hemisphere	(1, 21)	—	4.40*
Midline electrodes			
Condition	(1, 21)	6.20*	—
Condition × Electrode	(3, 36)	—	5.23*

Note. Analyses at trigger 4. Only significant results and trends are presented. † $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

ANOVAs for plural verbs in the 100–300 ms time window yielded a significant Condition main effect at midline and lateral electrodes, and a Condition×Laterality interaction in lateral electrodes (see Table 7). This interaction means that the negativity was strong at medial electrodes, but only marginally significant at more lateral electrodes. Given that the early negativity seemed most prominent over left-frontal electrodes (especially F3), the lack of interactions involving factors Hemisphere or Anteriority was somewhat surprising.

However, this was due to the fact that (a) the negativity was stronger at medial than lateral electrodes over *both* hemispheres, and (b) at posterior electrodes, the negativity was almost equally strong over both hemispheres (suggesting a second and more posterior N400-like negativity near the midline). An ANOVA in the P600 time-window (500–900 ms) yielded significant interactions of CONDITION×ELECTRODE at midline, and CONDITION× ANTERIORITY as well as CONDITION×HEMISPHERE at lateral electrodes. These interactions point to a posterior P600 co-occurring with an ongoing left-frontal negativity that gains strength once the P600 dissipates. In fact, between 800 and 1100 ms we found a significant CONDITION effect at F3 ($p < 0.02$) and Fz ($p < 0.03$), but not at more posterior electrodes. This pattern of an early frontal negativity and its reoccurrence after an intervening positivity is reminiscent of that previously described for various syntactic violations in auditory ERP studies (Steinhauer and Drury, 2012), suggesting a sustained frontal negativity and a temporarily overlapping P600. We will return to this below.

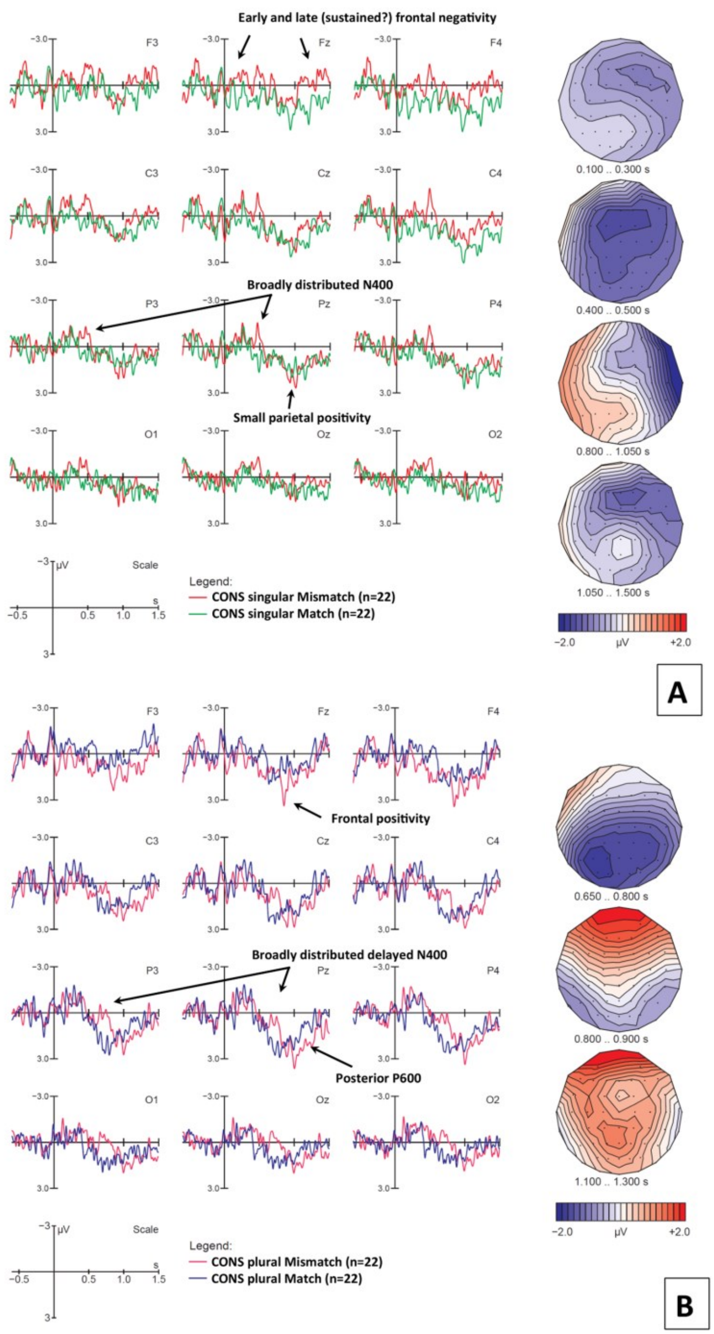
4.3.4.2. ERPs for consonant-final verb conditions at Trigger 4

While liaison verbs phonologically disambiguated number at verb onset, consonant verbs provided number information on the verb-final “morpheme” consonant. Due to this difference, one would expect mismatch effects to occur somewhat later than for liaison verbs. As shown in Figure 6A, for mismatch CONS *singular* verbs, the most prominent difference between match and mismatch conditions was a broadly distributed, slightly right-lateralized negativity in the N400 time window (400–500 ms after verb onset), which does not seem to be followed by a clear positivity in the P600 time-window. Note however that at anterior electrodes the N400 is both preceded and followed by a negativity starting around 100 ms, which seems to end around 600 ms and re-occur around 1000 ms. This pattern could, once again, reflect temporary ERP-component overlap, namely an early but sustained negativity with a frontal maximum (from 100–1500 ms), which is superimposed first by a parietal N400 that temporarily results in a more posterior scalp distribution (from 400–500 ms) and then by a left-lateralized and posterior positivity (from 800–1000 ms) that temporarily cancels out the negativity at most electrodes (especially over the left hemisphere), until the frontal negativity re-emerges. The assumption that the early (100–300 ms) and late negativity (1050–1500 ms) may reflect the same ongoing ERP component is supported by their similar scalp distribution (see first and last voltage maps in Figure 6A).

To test this assumption statistically, we ran ANOVAs directly comparing the two time windows (i.e., including the additional factor TIMEWINDOW). As expected, all significant effects involving the factor CONDITION were found to display the same scalp distribution in both time windows (100–300 ms and 1050–1500 ms, respectively), i.e., they did not interact with TIMEWINDOW. At midline electrodes, we found a CONDITION×ELECTRODE interaction (F(3, 63) 3.49, p = 0.04), reflecting a frontal negativity (in Fz only, F(1, 21) 5.59, p = 0.03), whereas lateral electrodes showed a main CONDITION effect (F(1, 21) 4.96, p = 0.04). In contrast, for the N400 between 400 and 500 ms, the ANOVA yielded significant CONDITION effects at midline and lateral electrodes, as well as a CONDITION×LATERALITY interaction at lateral electrodes (see Table 8). This interaction reflects a main CONDITION effect at medial electrodes. As a whole, this broadly distributed pattern along the midline strongly suggests the presence of a second (more posterior) negativity in addition to the ongoing frontal one. Lastly, in the P600 time window (800–1050 ms), we observe a significant CONDITION×HEMISPHERE interaction along with higher-order interactions involving CONDITION, HEMISPHERE, ANTERIORITY, and LATERALITY at lateral electrodes, and no effect at the midline. These interactions reflect a right-lateralized (and somewhat anterior) negativity, and a left-lateralized (somewhat posterior) positivity that largely cancel each other out at the midline (see third voltage map in Figure 6A).

Figure 6

ERP effects for number mismatches at consonant-final verbs (A) for singular and (B) for plural verbs.



Note. Displayed are grand-average ERPs at midline and lateral electrodes as well as voltage maps illustrating the difference waves, for all participants, time-locked to the onset of the critical verb using a baseline of -600 to 0 ms. The vertical bar marks the onset of the critical verb. (A) Compared to the match condition (green lines), singular mismatches (red lines) show an early sustained negativity at frontal electrodes (100-1500 ms; cf. Voltage maps 1 and 4), an additional N400 (400-500 ms), and an intermediate time window during which a right-anterior negativity and a left-posterior negativity seem to cancel each other out along the midline (800-1050 ms). (B) Compared to the match condition (blue lines), plural mismatches (magenta lines) show an N400-like negativity (650-800 ms), followed by a frontal positivity (800-900 ms) and a posterior P600 (1100-1300 ms).

Table 8

Global repeated measures ANOVAs for consonant-final singular verbs in neutral contexts at time-windows of interest.

		(N400)	(P600)	Negativity
	df	400–500	800–1050	1050–1500
LATERAL ELECTRODES				
CONDITION	(1, 21)	6.88*	—	4.72*
CONDITION × LATERALITY	(1, 21)	6.36*	—	—
MEDIAL: CONDITION	(1, 21)	9.30**	—	—
CONDITION × HEMISPHERE	(1, 21)	—	5.68*	—
RIGHT HEM: CONDITON	(1, 21)	—	6.70*	—
CONDITION × HEM × ANTERIORITY	(1, 21)	—	3.56†	—
CONDITION × HEM × LATERALITY	(1, 21)	—	6.55*	—
RIGHT HEM: CONDITION	(1, 21)	—	6.67*	—
LATERAL: CONDITION	(1, 21)	—	5.85*	—
CONDITION × HEM × LAT × ANT	(2, 42)	—	6.72**	—
LEFT HEM: CON × LAT × ANT	(2, 42)	—	4.08*	—
LEFT HEM: FRONT: CON × LAT	(1, 21)	—	5.53*	—
MIDLINE ELECTRODES				
CONDITION	(1, 21)	7.78**	—	—
CONDITION × ELECTRODE	(3, 36)	—	—	3.66*
FZ: CONDITION	(1, 21)	—	—	5.54*

Note. Analyses at trigger 4. Only significant results and trends are presented. Ant = Anteriority, Cent = Central, Con = Condition, Front = Frontal, Hem = Hemisphere, Lat = Laterality, Post = Posterior, † $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

For CONS *plural* verbs (depicted in Figure 6B, statistics in Table 9) we observe a number mismatch effect reflected by a more delayed N400 than in singular contrasts (650–800 ms after verb onset), followed by a frontal P3a-like positivity (800–900 ms) and a late posterior one (1100–1300 ms). We ran an ANOVA for plural CONS verbs in the later N400 time-window (650–800 ms). This yielded a significant CONDITION main effect at midline and a

CONDITION×ANTERIORITY interaction at lateral electrodes. Decomposition of this interaction revealed a main CONDITION effect at both central and posterior electrodes. An ANOVA in the 800–900 ms time-window yielded a CONDITION× ELECTRODE interaction at midline, and a CONDITION×ANTERIORITY interaction at lateral electrodes. These interactions reflect a significant frontal positivity (main effects of CONDITION at Fz), and a corresponding trend at anterior lateral electrodes. Finally, a main effect of CONDITION was found in the late P600 (1100–1300 ms) time-window, but only in posterior electrodes. No other main effects or interactions were found.

Table 9

Global repeated measures ANOVAs for consonant-final plural verbs at time-windows of interest.

		(N400)	(Late N400)	(P600, frontal)	(P600, posterior)
	df	400–500	650–800	800–900	1100–1300
LATERAL ELECTRODES					
CONDITION	(1, 21)	—	—	—	—
CONDITION × ANTERIORITY	(2, 42)	—	4.36*	5.50*	—
ANTERIOR: CONDITION	(1, 21)	—	—	3.25†	—
CENTRAL: CONDITION	(1, 21)	—	5.45*	—	—
POSTERIOR: CONDITION	(1, 21)	—	8.41**	—	10.05**
MIDLINE ELECTRODES					
CONDITION	(1, 21)	—	5.47*	—	—
CONDITION × ELECTRODE	(3, 36)	—	2.88†	4.62*	—
FZ: CONDITION	(1, 21)	—	—	4.50*	—

Note. Analyses at trigger 4. Only significant results and trends are presented. † $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

4.3.4.3. Discussion for number mismatches on verbs

Whereas cross-modal lexico-semantic mismatches have been shown to elicit N400s in a number of previous studies, number mismatches between visual and auditory input have not been studied so far. Given that our paradigm used grammatical sentences it was unclear whether our number mismatches would elicit ERP profiles typical for “grammatical” agreement errors, i.e., LAN/N400s and P600s. Number disambiguation in neutral contexts only became available on the verbs. Not unlike mismatch effects at sentence onset, ERPs at verb onset elicited biphasic

(N400-P600) profiles in three out of four contrasts. As expected, component latency was influenced by the availability of disambiguating number information (earlier for verb-initial liaisons than for verb-final consonants, and earlier for shorter singular than for longer plural CONS verbs). In addition, two conditions (LIAIS plural and CONS singular) displayed sustained anterior negativities, resulting in complex patterns of overlapping ERP components. In contrast, singular LIAIS mismatches did not display any systematic ERP effects at all.

For LIAIS verbs, we first discuss the lack of ERP components for the singular condition before turning to effects found in the plural.

4.3.4.4. Discussion for singular liaison verbs

The absence of ERP effects in the singular LIAIS condition corresponds to relatively poor behavioral performance in that particular condition, i.e., sentences with neutral contexts (e.g., *'For dessert, she likes…'* concurrently with an image illustrating two girls). The different ERP mismatch effects for singular vs. plural sentences with neutral contexts in LIAIS verbs may therefore reflect these difficulties. Note that we cannot explain these effects by appealing to differences between commission and omission, nor plural vs. singular forms (singular being the default), since CONS singular forms *did* elicit ERP components. Similarly rule strength or predictability would promote better perception of differences in liaison, as this process is obligatory in French, and also reliably occurs in determiner-noun contexts. We explore phonological salience, truth-value interpretations assigned to sentences, and sociolinguistic variability as explanations for these results.

Phonological salience (or perceptual salience) refers to the ease with which we can hear or perceive a given structure (Goldschneider and DeKeyser, 2005). Applied to our materials, we can expect that arriving at an accurate sentence interpretation is facilitated by overt phonological cues for number. We used an overt cue for number with LIAIS verbs, which in the plural is arguably more salient – due to the presence of a /z/ – than in the singular without a /z/. It seems very unlikely that a participant – after hearing *elles aiment* [ɛlz̲ɛm] – would be willing to deny the cue's presence and assume she may have hallucinated, just because the picture only shows a single potential subject. However, if the same participant sees a picture with two girls and hears singular forms such as *elle aime* [ɛlɛm], it seems possible to conclude to having misperceived liaison. Similar differences between the presence versus absence of phonological (and visual) evidence have been found for prosodic boundaries and commas (leading to the 'Boundary Deletion Hypothesis', cf. Steinhauer and Friederici, 2001; Pauker et al., 2011). Phonological salience thus seems to provide a plausible explanation for the absence of ERP mismatch effects for singular sentences with neutral contexts. However, it does not account for all of our data, as singular CONS mismatches (which were also marked by a non-salient cue) did in fact elicit ERP responses.

Alternatively, the null result for LIAIS singular mismatches might be due to their enhanced acceptability, based on truth-values. Acceptability assigned to our sentences can be either logically or pragmatically motivated. For example, the sentence *Some triangles have three edges* is logically true, but under-informative and pragmatically odd. Similarly, when presented with an image of two girls eating chocolate mousse, describing the picture with "*She likes ...*" is also logically true, but pragmatically odd. The ERP literature suggests that people differ in their bias towards logical versus pragmatic processing (e.g., Barbet and Thierry, 2016). If some of our participants were biased towards logical processing, we would expect reduced or absent mismatch effects for neutral singular mismatching sentences. Crucially, however, even though one could argue that a lacking mismatch effect due to logical processing biases should be limited to singular sentences, there is no reason why it should be limited to sentences that are disambiguated by LIAIS verbs. That is, number mismatches disambiguated by CONS verbs would be subject to the same logic, but they did elicit clear ERP mismatch effects.

Yet another way of explaining the absence of ERPs for LIAIS singular mismatches comes from sociolinguistics. According to Prof. Julie Auger at Indiana University (personal communication), *elles* 'she.PLUR' does not exist in informal Québec French, due to a process of neutralisation (i.e., masculine and feminine plural pronoun clitics have become indistinguishable). Both are pronounced [i] before a consonant and [j] before a vowel (e.g., *les filles/les garcons y'aiment* 'The girls/the boys, they like' are equally grammatical), although there is some variability between dialects. Two corpora from French monolingual speakers in Quebec City and bilingual speakers in Ottawa-Hull reveal few uses of *elles*, and omission or replacement of *elles* by *ils* 'they.MASC' in addition to /l/-deletion (i.e., /il/ or /ilz/ pronounced [i], [iz] or [j], but rarely [ɛl/z] or [ɪl/z] the standard forms for plural) (Bourget, 1987; Poplack and Walker, 1986). The [j], being a semi-vowel, is licit before a vowel-onset verb and no additional liaison is necessary, and could in fact block liaison, since the verb onset is filled. Thus, perception of a subject-verb agreement error in liaison might be less systematic in singular conditions due to loss, or variability, of this grammatical feature, an interpretation that is coherent with our behavioural data where only these forms showed lower accuracy rates. We do not know of a psycholinguistic study that directly investigates liaison processing in Québec French, and so this interesting account remains somewhat speculative. While it appears to best explain our ERP null result for LIAIS singular mismatches (and is not applicable to CONS verbs), we should recall that participants still recognized the mismatches more than 85 % of the time. We suggest that the absence of consistent ERP effects with LIAIS singular verbs reflects increased variability in processing strategies across participants, which may very well be influenced by sociolinguistic variability. As reflected by later sentence-wrap-up effects (see Supplementary materials manuscript 2, section 4.7), in some cases error processing might also have been delayed.

4.3.4.5. Discussion for plural liaison verbs

The early-onset and sustained frontal negativity for plural mismatches resembles a classic morphosyntactic (dis-)agreement effect in auditory studies, possibly corresponding to more short-lived LAN-like effects in reading studies (Hasting and Kotz, 2008; Steinhauer and Drury, 2012). The extremely short onset latency of this effect, around 100 ms in our data, may be slightly overestimated due to possible co-articulation prior to the verb onset trigger (Trigger 4 in Figure 1) and the presence of the phoneme /z/ indexing a plural pronoun preceding it. As with Hasting and Kotz (2008), this is another illustration that morphosyntax-related processing difficulties that are clearly *not* driven by phrase structure violations can elicit this type of negativity (contra Friederici, 2002, 2011). Another similarity with Hasting and Kotz (2008), as well as many other auditory studies, is our finding of a complex pattern of overlapping ERP components (as discussed in Steinhauer and Drury, 2012). That is, sustained negativities are often superimposed by posterior P600 effects leading to a temporary mutual cancellation of components in at least certain electrodes. In our particular case, the negativity's scalp distribution in the early 100–300 ms time-window points to an even more complex pattern, as the P600 (500–800 ms) seems to be preceded by an additional, more posterior (N400-like) negativity from 100–300 ms that also overlaps with the frontal negativity. In our opinion, this is what explains the rather broad distribution of negativities in this time-window as reflected by statistical analyses, whereas the last portion of the "re-emerging" frontal negativity was limited to left-frontal electrode sites.

As with mismatch effects at sentence onset, the N400 effect may primarily indicate a lexical/phonological mismatch with what was predicted based on the picture. That is, participants saw a single person (e.g., one girl eating, thus predicting *elle* [ɛl], i.e., 'she') but heard sentences such as *Au dessert, <u>elles</u> aiment* ... ('For dessert, <u>they</u> like …'). Importantly, at least initially this mismatch is compatible with a number of interpretations. First, it is possible that the perceived mismatch included both the pronoun and the verb (*elles aiment* 'they.FEM.PLUR like' instead of *elle aime* 'she.FEM.SING likes'). This implies that the auditorily presented sentence as a whole was processed as a grammatical plural sentence, and the pronoun+verb as a whole mismatched across modalities. The first mismatching cue was provided by the pronoun at verb onset (liaison) and elicited an N400, as with NP contexts at sentence onset. The subsequent P600 was also triggered by the pronoun+verb and either reflected conflict monitoring and mismatch resolution or task-relevant categorization of a mismatching trial, or both. In this scenario, it is also possible that participants considered a *generic* interpretation. That is, 'they (i.e., girls) like chocolate mousse' is an assertion that, in principle, could be illustrated with one single girl. As in English, French generic expressions are realized in the plural. However, for a generic (acceptable) interpretation we would predict a higher acceptability rate (which we did not find) and not expect a P600 (which we did find). Secondly, it is possible that the visual presentation of a single person activated a very strong expectation for a singular sentence. Knowing that incoming sentences were always supposed to describe the pictures, all spoken information up to phoneme /z/ at the liaison (including the entire context and most of the pronoun [ɛl]) was compatible with a singular interpretation, and it is conceivable that the longer the ambiguity lasted, the more this singular interpretation was strengthened. This expectation of a singular sentence may have led to two processing strategies that are both distinct from the first one discussed above: One is that only the pronoun, but not the verb, was processed as a plural form. Recall that liaison verbs were phonologically indistinguishable between singular and plural, i.e., *aime/nt* [ɛm]. So hearing *elles aiment'* [ɛlzɛm] could have been interpreted as *elles *aime*, 'they likes', a classical morphosyntactic agreement violation. In this scenario, the P600 would reflect some process of reanalysis towards a singular interpretation. The other possibility assumes that the initial expectation of a singular sentence was so strong that it led participants to temporarily mis-parse the incoming speech signal. Instead of interpreting /z/ as the pronoun plural marker (*elles* [ɛlz] + *aime(nt)* [ɛm]) they may have interpreted it as a verb-initial phoneme (i.e., *elle* [ɛl] + *zaime(nt)* [zɛm]). This latter scenario is a possibility, as certain properties of French may have supported this. For instance, pronouns do not normally carry stress and are cliticized with the next content word to form one prosodic word where the content word carries word-final stress. Moreover, according to the "maximal onset principle" (Selkirk, 1981), the plural pronoun marker /z/ is syllabified into the verb's first syllable, as [ɛl.**zɛm**] and not [ɛlz.**ɛm**] (bold font indicates stress). This is the same pattern one would expect for a singular utterance (i.e., *elle zaime*). Importantly, even though the verb *zaimer* does not exist in French, there are a number of French verbs that do start with /z/ (e.g., *zigonner* 'to dally', *zigouiller* 'to kill', *zigzaguer* 'to zigzag', *zézayer* 'to lisp', *zyeuter* 'to observe intently', *zébrer* 'to decorate with stripes'). In other words, given the large number of different verbs used in our study (without any within-subject repetition), it is conceivable that in the LIAIS plural condition participants might have checked their lexicon for a verb that starts with /z/. We propose that ambiguity complexity in this particular condition may have elicited the additional sustained negativity, possibly reflecting evaluation of multiple options.

4.3.4.6. Discussion for singular consonant-final verbs

We will now turn to number mismatches on consonant-final verbs. The singular CONS mismatch condition with neutral contexts again elicited three components: a sustained anterior negativity (AN), an N400 and a small slightly left-lateralized P600. This pattern resembles that found in plural LIAIS mismatches with, however, a reduced P600. The later onset for the N400 as compared with LIAIS verbs can be straightforwardly explained by the later appearance of disambiguating information in the CONS condition's sound-streams. Interestingly the AN does not differ in distribution between early and late time-windows. According to Steinhauer and Drury (2012), this is one way of demonstrating that two negativities are likely early and late portions of the *same* (ongoing) ERP component. In the intervening time-windows, it is first superimposed by an N400 and then canceled out by a P600, which themselves may have overlapped and canceled each other out to some extent (explaining the absence of either effect between 500 and 800 ms). In contrast to both sentence onset and LIAIS verb conditions, here number ambiguities lasted until the verb-final consonant. That is, when participants saw a picture of two lions roaring and heard *En soirée il <u>rugit</u>* [ilʁyʒ**i**] *dans la savane* 'In the evening he roars in the savannah', only the lack of the verb-final consonant [s] (*rugissent* [ilʁyʒɪ**s**]) indicated a mismatch. Importantly, as the singular and plural pronouns *il* and *ils* are homophonous ([il]), we assume that the pronoun was initially processed as a plural (as suggested by the picture). Thus, one interpretation of what happened at the disambiguation point is that participants interpreted the auditory input as *ils *rugit*, ('he. PLURAL roar. SINGULAR'), which corresponds to a classical oral-language agreement violation. As before, the N400 would reflect a lexical-phonological mismatch, and the P600 would be associated with both with categorization of this sentence as a mismatch and a potential attempt to revise its structure. Recall however, that (a) phonologically, the absence (omission) of a verb-final consonant is not very salient, and (b) participants were strongly biased towards a plural interpretation. Therefore, it is conceivable that participants were not entirely sure if the perceived mismatch was real or if they had simply missed an actually present consonant. Similar temporary confusions based on strong predictions are known from e.g., Itzhak et al (2010) who demonstrated that listeners perceive a prosodic boundary in absence of any acoustic markers, if both lexical information and syntactic structure strongly predict it. Moreover, and only in the CONS singular condition, it is possible that participants initially parsed the subsequent preposition's word-initial consonant as a verb-final plural marker. In our example (*il(s) rugit* [ilʁyʒ**i**] *dans ...)*. Misinterpreting the /d/ of *dans* as a plural marker would result in [ilʁyʒi**d**], which could – in principle – be interpreted as a plural verb form (i.e., *ils rugident*). However, in the singular, the stem-final vowel is stressed due to the absence of a word-final coda (compare *ils rugissent* [ilʁyʒ**ɪs**]), and is a strong cue to word structure. At this point, participants would need to check this verb's stem forms in their mental lexicon and verify which one is legal in the plural. We believe that the complexity involved in this ambiguity is the reason why we find, once again, a sustained frontal negativity, resembling the LIAIS plural condition. As in previous conditions, we interpret the N400 as a reflection of an initial lexical-phonological mismatch, and the P600 as an attempt to resolve its structural consequences. The fact that the frontal negativity lasted beyond the P600 duration (as in LIAIS plurals and previous auditory agreement studies, e.g., Hasting and Kotz, 2008) suggests that the P600 does not always reflect the final stage of evaluation processes. One particularity of the CONS singular mismatch pattern was that the P600 itself did not reach statistical significance. Several previous studies have refrained from interpreting similar findings (e.g., Hasting and Kotz, 2008; Ye et al., 2006), but Steinhauer and Drury (2012) have argued that in the presence of ongoing negativities, existence of a P600 can be inferred if this negativity is temporarily cancelled out during the P600 time window (and at plausible electrode sites) and then re-emerges. We will come back to this point below.

4.3.4.7. Discussion for plural consonant-final verbs

Unlike singular CONS verbs, mismatches with plural CONS verbs elicited only a posterior N400 followed by a large P600, but no AN. As expected (see above), both components emerged slightly later than in the singular condition (due to the longer plural form duration). In many ways the plural condition resembles the singular one, however, the mismatching information is (a) phonologically salient and (b) an unambiguous plural verb marker. Thus, once plural information has been encountered, there can be no doubt that the verb is incompatible with an initial assumption of a singular pronoun (akin to a garden path sentence). In our example, the most likely lexical representation would be *En soirée il *<u>rugissent</u>* [ilʁyʒı**s**] – a classical case of morphosyntactic number disagreement. In fact, we believe that – of all number mismatch conditions in our study – this condition is closest to a traditional oral-language agreement violation. As both the presence and the nature of this mismatch are extremely obvious, both the N400 and the P600 were found to be strong and consistent, while no AN reflecting effortful evaluation of a more ambiguous scenario was elicited.

4.3.4.8. Sentence-final negativities and wrap-up effects

A subset of number-mismatch conditions (see Supplementary Materials), as well as the lexico-semantic condition, elicited a late posterior negativity at sentence end (1700–2000 ms), which we interpret as potential "sentence wrap-up" effects for both types of error. In contrast to positive waveforms that tend to occur in sentence-final positions of correct sentences, negativities are typically associated with preceding linguistic anomalies and may reflect additional processing load involved in reconsidering the anomaly and integrating the entire sentence. A recent study from our lab on conceptual and logical semantic anomalies also showed that sentence final N400-like "wrap-up" effects are common, irrespective of the type of linguistic violation occurring in mid-sentence positions and of whether these elicited local N400s or P600s (Bokhari, 2015). Recently, Stowe and colleagues (2018) have raised the question of whether "sentence wrap-up effect" is an appropriate label for these negativities given the link to anomalies; these authors suspect that task requirements may also play a role in eliciting them. "Anomaly-related sentence-final negativity" may thus be a more neutral term to characterize these ERP effects.

4.4. Discussion for all conditions

The present study used ERPs to investigate whether visual-auditory mismatches between a picture and a perfectly grammatical spoken sentence would elicit similar brain responses as typically seen for *within*-sentence linguistic anomalies. We included both cross-modal semantic mismatches, realized on verbs, and number mismatches (singular vs. plural) that occurred at different sentence positions using a range of linguistic number markers in spoken French (determiners, liaison, and verb-final consonants). Analyses also contrasted potential differences between singular and plural mismatches. Overall, our data demonstrate that cross-modal mismatches result in ERP profiles known from the literature for linguistic anomalies, and seem to distinguish between mismatches that can be described as purely conceptual-semantic and those that can be viewed as concerning grammar.

4.4.1. N400s, P600s and ANs – Evidence for agreement violations?

Returning to our initial research questions, our data have demonstrated that (a) cross-modal semantic mismatches realized on verbs elicit typical N400s and that (b) participants use the first available linguistic cues to detect number mismatches between a picture and a spoken sentence. Whether the ERP components found for cross-modal number mismatches are indistinguishable from those typically observed for "purely linguistic" within-sentence agreement violations, is less clear. On the one hand, all components we observed for number mismatches are within the range of ERP effects previously observed for morphosyntactic agreement violations. On the other hand, Molinaro et al. (2011) reported that previous studies on number agreement violations have typically found LANs and P600s. While most of our negativities preceding the P600s did show a LAN-like frontal distributions, sometimes even with a left-lateralized prominence, statistical evidence usually pointed to a broadly distributed negativity compatible with an N400. Moreover, clearer evidence for left-anterior negativities (i.e., in LIAIS plural verbs) could be attributed to an early-onset sustained negativity at left frontal electrodes (e.g., F3). Overall, we believe our data are more compatible with an N400-P600 profile than with a LAN-P600 one. However, most previous ERP studies on number (dis-)agreement have focused on effects within NPs (determiner-adjective-noun) in the written modality. It is still controversial to what extent LANs (especially in reading studies) result from component overlap between N400s and P600s (e.g., Tanner and Van Hell, 2014). However, our data do provide more evidence showing that early-onset sustained negativities in mismatch studies can show a clear left-anterior distribution that cannot be explained by component overlap. Since LANs in reading studies tend to have latencies and durations comparable to N400s (i.e., 300–500 ms), we are increasingly less convinced that sustained (left-)anterior negativities in auditory studies (e.g., Hasting and Kotz, 2008; van den Brink and Hagoort, 2004) are analogous to LAN components in reading studies. For our current data, we suggest that sustained negativities may index a continued evaluation of more complex cases of ambiguity resolution. The N400s we found virtually in all number mismatch conditions are rather difficult to interpret with confidence, as various accounts would predict N400-like components, including for standard morphosyntactic violations involving predictable inflectional morphemes (e.g., Tanner, 2015; BSS2018), truth-value related approaches (Bokhari, 2015), and phonological mismatch accounts (Connolly and Phillips, 2004). Molinaro and colleagues (2011) have argued that phonotactics involved in agreement processes might demote grammatical processing (reflected by LANs) towards a lexical one (reflected by N400s). Our CONS verbs had a variety of final consonants (9 different consonants over our 60 verbs). These consonant changes do not follow systematic morphological rules. They are sometimes described as consonant deletion rules from the plural to the singular (Paradis and El Fenne, 1995). However, since singular forms are the default (and are acquired first), Royle (2011) argues against this approach and proposes rather that consonant-alternating forms in French are lexicalized (her research focused on adjectives, but the same logic can also be applied to verbs). This could promote use of lexical rather than grammatical processing when checking agreement, and thus explain N400 effects observed in plural conditions.

4.5. Conclusion

With the aim of testing whether cross-modal mismatches between pictures and *grammatical* sentences would elicit similar ERP components to those in the literature on linguistic anomalies, we developed an experiment with auditory-visual sentence-picture matching paradigms and an acceptability judgment task in French. We investigated neurocognitive mechanisms underlying lexico-conceptual semantics and grammatical number processing. This is the first study to test three different linguistic cues for number mismatches at different sentence positions. Our results demonstrated that native French speakers reliably exhibit N400 components in response to cross-modal verb-action mismatches, comparable to previous effects found for noun-object mismatches. Auditory-visual number mismatches usually elicited a biphasic N400-P600 (in some cases superimposing a sustained AN), and our context manipulation demonstrated that participants use the first available sentence cue to disambiguate structures. ERP effects at sentence onset and on the verb suggest that participants immediately tracked mismatches between modalities as soon as conflicting information became available, and that these mismatches were processed in a way that is not fundamentally different from purely linguistic within-sentence agreement violations.

Our paradigm is exciting for a number of reasons, one being that we used grammatical sentences to induce "agreement error" processing and elicited well-known ERP components. This approach has the advantage of being more ecologically valid than error-based paradigms, as it resembles more closely the mostly error-free speech we are exposed to daily. Having developed this experiment for younger populations, we are confident that our approach will reveal, in children, what types of information are being used at which point in the speech stream to disambiguate information. This type of paradigm also has potential for the study of developmental language disorder as well as second-language learning, as is the visual-world paradigm used in eye-tracking studies (e.g., Hopp and Lemmerth, 2018).

We can anticipate future directions of inquiry from this initial study of verb-based visual-auditory mismatches. As we have seen, not all incongruent number mismatch conditions elicited strong P600s despite the fact that we used a judgment task, which promotes this component. The

N400 component seemed to be a more reliable reflection of our mismatch errors. This might in part be due to the fact that we did not use ungrammatical sentences as input, reducing error-detection based strategies that could have been used in most studies that find the LAN or the P600. Our robust N400s instead of LANs (or ANs), and less robust P600s for mismatches, might be the result of our sentences' *grammatical* status.

As we have appealed to sociolinguistics to explain some of our results, it appears interesting to pursue sociolinguistic studies using ERPs. This combination of domains has rarely been explored and we can identify straightforward implementations, as in second language acquisition research, to study variability in grammars within geographically constrained but linguistically diverse speakers of the same language. Paying attention to how a speaker implements a particular linguistic rule has strong potential to help us better understand the neurocognitive underpinnings of within-group variability in language processing.

In conclusion, our study provides a significant contribution to the field of cognitive neuroscience of language by providing high-quality evidence regarding the generalizability of ERP profiles across modalities and languages. This study extends lexico-semantic mismatches to the domain of verbs, provides insight into context effects and early detection of mismatches, establishes ERP patterns for different types of morpho-phonological and morpho-syntactic cues for number mismatch processing, and demonstrates that even grammatical sentences can elicit ERP patterns associated with "error" processing.

4.6. References

Angrilli, A., Penolazzi, B., Vespignani, F., De Vincenzi, M., Job, R., Ciccarelli, L., et al. (2002). Cortical brain responses to semantic incongruity and syntactic violation in Italian language: an event-related potential study. *Neurosci. Lett.* 322, 5–8. doi:10.1016/S0304-3940(01)02528-9.

Balconi, M., and Pozzoli, U. (2005). Comprehending semantic and grammatical violations in Italian. N400 and P600 comparison with visual and auditory stimuli. *J. Psycholinguist. Res.* 34, 71–98. doi:10.1007/s10936-005-3633-6.

Barber, H., and Carreiras, M. (2005). Grammatical gender and number agreement in Spanish: An ERP comparison. *J. Cogn. Neurosci.* 17, 137–153. doi:10.1162/0898929052880101.

Barbet, C., and Thierry, G. (2016). Some alternatives? Event-related potential investigation of literal and pragmatic interpretations of some presented in isolation. *Front. Psychol.* 7, 1479. doi:10.3389/fpsyg.2016.01479.

Bokhari, F. S. (2015). Investigating the neurocognitive mechanisms underlying truth-conditional and logical semantic aspects of sentence processing: An event-related brain potential study (Master's Thesis, McGill University, 2015). Available at: http://digitool.library.mcgill.ca/webclient/StreamGate?folder_id=0&dvs=1552416877800~191&#search=%22bokhari%22&usePid1=true&usePid2=true

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2018, November 20). Towards a neurobiologically plausible model of language-related, negative event-related potentials. *OSF Preprints.* doi:10.31219/osf.io/aynsp

Bourget, M.-J. (1987). "Variation phonétique dans l'emploi des pronoms de troisième personne en français," in *Tendances actuelles de la recherche sur la langue parlee [Current Trends in Research on Oral Language]*, eds. J. Auger, R. Grenier, R. Lapalme, J.-F. Montreuil, and P. Whitmore (Québec, Canada: Université Laval), 129–140.

Brink, D. V. D., & Hagoort, P. (2004). The influence of semantic and syntactic context constraints on lexical selection and integration in spoken-word comprehension as revealed by ERPs. *Journal of Cognitive Neuroscience*, *16*(6), 1068-1084. doi: 10.1162/0898929041502670

Brown, C., and Hagoort, P. (1993). The processing nature of the N400: Evidence from masked priming. *J. Cogn. Neurosci.* 5, 34–44. doi:10.1162/jocn.1993.5.1.34.

Coulson, S., King, J. W., and Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Lang. Cogn. Process.* 13, 21–58. doi:10.1080/016909698386582.

DeLong, K. A., Urbach, T. P., and Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nat. Neurosci.* 8, 1117. doi:10.1038/nn1504.

Dube, S., Kung, C., Peter, V., Brock, J., and Demuth, K. (2016). Effects of type of agreement violation and utterance position on the auditory processing of subject-verb agreement: an ERP study. *Front. Psychol.* 7, 1276. doi:10.3389/fpsyg.2016.01276.

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

Foucart, A., and Frenck-Mestre, C. (2011). Grammatical gender processing in L2: Electrophysiological evidence of the effect of L1–L2 syntactic similarity. *Biling. Lang. Cogn.* 14, 379–399. doi:10.1017/S136672891000012X.

Foucart, A., and Frenck-Mestre, C. (2012). Can late L2 learners acquire new grammatical features? Evidence from ERPs and eye-tracking. *J. Mem. Lang.* 66, 226–248. doi:10.1016/j.jml.2011.07.007.

Frenck-Mestre, C., Osterhout, L., McLaughlin, J., and Foucart, A. (2008). The effect of phonological realization of inflectional morphology on verbal agreement in French: Evidence from ERPs. *Acta Psychol. (Amst.)* 128, 528–536. doi:10.1016/j.actpsy.2007.12.007.

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends Cogn. Sci.* 6, 78–84.

Friederici, A. D. (2011). The brain basis of language processing: from structure to function. *Physiol. Rev.* 91, 1357–1392. doi:10.1152/physrev.00006.2011.

Friederici, A. D., Mecklinger, A., Spencer, K. M., Steinhauer, K., and Donchin, E. (2001). Syntactic parsing preferences and their on-line revisions: A spatio-temporal analysis of event-related brain potentials. *Cogn. Brain Res.* 11, 305–323. doi:10.1016/S0926-6410(00)00065-3.

Friederici, A. D., Pfeifer, E., and Hahne, A. (1993). Event-related brain potentials during natural speech processing: Effects of semantic, morphological and syntactic violations. *Cogn. Brain Res.* 1, 183–192. doi:10.1016/0926-6410(93)90026-2.

Friedrich, M., and Friederici, A. D. (2004). N400-like semantic incongruity effect in 19-month-olds: processing known words in picture contexts. *J. Cogn. Neurosci.* 16, 1465–1477. doi:10.1162/0898929042304705.

Goldschneider, J. M., and DeKeyser, R. M. (2005). Explaining the “Natural Order of L2 Morpheme Acquisition” in English: A meta-analysis of multiple determinants. *Lang. Learn.* 55, 27–77. doi:10.1111/j.0023-8333.2005.00295.x.

Haebig, E., Weber, C., Leonard, L. B., Deevy, P., and Tomblin, J. B. (2017). Neural patterns elicited by sentence processing uniquely characterize typical development, SLI recovery, and SLI persistence. *J. Neurodev. Disord.* 9, 22. doi:0.1186/s11689-017-9201-1.

Hagoort, P. (2003). Interplay between syntax and semantics during sentence comprehension: ERP effects of combining syntactic and semantic violations. *J. Cogn. Neurosci.* 15, 883–899. doi:10.1162/089892903322370807.

Hagoort, P., and Brown, C. M. (2000). ERP effects of listening to speech compared to reading: the P600/SPS to syntactic violations in spoken sentences and rapid serial visual presentation. *Neuropsychologia* 38, 1531–1549. doi:10.1016/S0028-3932(00)00053-1.

Hahne, A., and Friederici, A. D. (1999). Electrophysiological evidence for two steps in syntactic analysis: Early automatic and late controlled processes. *J. Cogn. Neurosci.* 11, 194–205. doi:10.1162/089892999563328.

Hasting, A. S., and Kotz, S. A. (2008). Speeding up syntax: On the relative timing and automaticity of local phrase structure and morphosyntactic processing as reflected in event-related brain potentials. *J. Cogn. Neurosci.* 20, 1207–1219. doi:10.1162/jocn.2008.20083.

Holcomb, P. J. (1993). Semantic priming and stimulus degradation: Implications for the role of the N400 in language processing. *Psychophysiology* 30, 47–61. doi:10.1111/j.1469-8986.1993.tb03204.x.

Holcomb, P. J., and Neville, H. J. (1990). Auditory and visual semantic priming in lexical decision: A comparison using event-related brain potentials. *Lang. Cogn. Process.* 5, 281–312. doi:10.1080/01690969008407065.

Hopp, H., & Lemmerth, N. (2018). Lexical and syntactic congruency in L2 predictive gender processing. *Studies in Second Language Acquisition, 40*, 171–199. doi:10.1017/S0272263116000437

Itzhak, I., Pauker, E., Drury, J.E., Baum, S.R. and Steinhauer, K., 2010. Event-related potentials show online influence of lexical biases on prosodic processing. *NeuroReport*, *21*, 8-13. doi: 10.1097/WNR.0b013e328330251d

Kolk, H. H., Chwilla, D. J., Van Herten, M., and Oor, P. J. (2003). Structure and limited capacity in verbal working memory: A study with event-related potentials. *Brain Lang.* 85, 1–36. doi:10.1016/S0093-934X(02)00548-5.

Kutas, M., and Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends Cogn. Sci.* 4, 463–470. doi:10.1016/S1364-6613(00)01560-6.

Kutas, M., and Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annu. Rev. Psychol.* 62, 621–647. doi:10.1146/annurev.psych.093008.131123.

Kutas, M., and Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science* 207, 203–205. doi:10.1126/science.7350657.

Kutas, M., and Hillyard, S. A. (1983). Event-related brain potentials to grammatical errors and semantic anomalies. *Mem. Cognit.* 11, 539–550. Available at: https://link.springer.com/article/10.3758%2FBF03196991.

Kutas, M., Van Petten, C. K., and Kluender, R. (2006). "Psycholinguistics electrified II (1994–2005)," in *Handbook of Psycholinguistics (Second Edition)*, eds. M. J. Traxler and M. A. Gernsbacher (Elsevier), 659–724.

Lau, E. F., Phillips, C., and Poeppel, D. (2008). A cortical network for semantics:(de) constructing the N400. *Nat. Rev. Neurosci.* 9, 920. doi:10.1038/nrn2532.

Lau, E., Stroud, C., Plesch, S., and Phillips, C. (2006). The role of structural prediction in rapid syntactic analysis. *Brain Lang.* 98, 74–88. doi:10.1016/j.bandl.2006.02.003.

Luck, S. J. (2014). *An introduction to the event-related potential technique*. 2nd ed. Cambridge, MA: MIT press.

Molinaro, N., Barber, H. A., and Carreiras, M. (2011a). Grammatical agreement processing in reading: ERP findings and future directions. *Cortex* 47, 908–930. doi:10.1016/j.cortex.2011.02.019.

Molinaro, N., Vespignani, F., Zamparelli, R., and Job, R. (2011b). Why brother and sister are not just siblings: Repair processes in agreement computation. *J. Mem. Lang.* 64, 211–232. doi:10.1016/j.jml.2010.12.002.

Morgan-Short, K., Sanz, C., Steinhauer, K., and Ullman, M. T. (2010). Second language acquisition of gender agreement in explicit and implicit training conditions: An event-related potential study. *Lang. Learn.* 60, 154–193. doi:10.1111/j.1467-9922.2009.00554.x.

Nevins, A., Dillon, B., Malhotra, S., & Phillips, C. (2007). The role of feature number and feature-type in processing Hindi verb agreement violations. *Brain Research, 1164*, 81–94 doi:10.1016/j.brainres.2007.05.058

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9, 97–113. doi:10.1016/0028-3932(71)90067-4.

Osterhout, L., and Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *J. Mem. Lang.* 31, 785–806. doi:10.1016/0749-596X(92)90039-Z.

Osterhout, L., and Holcomb, P. J. (1993). Event-related potentials and syntactic anomaly: Evidence of anomaly detection during the perception of continuous speech. *Lang. Cogn. Process.* 8, 413–437. doi:10.1080/01690969308407584.

Osterhout, L., and Mobley, L. A. (1995). Event-related brain potentials elicited by failure to agree. *J. Mem. Lang.* 34, 739–773. doi:10.1006/jmla.1995.1033.

Paradis, C., and El Fenne, F. (1995). French verbal inflection revisited: Constraints, repairs and floating consonants '? *Lingua* 95, 169–204. doi:10.1016/0024-3841(95)90105-1.

Pauker, E., Itzhak, I., Baum, S. R., and Steinhauer, K. (2011). Effects of cooperating and conflicting prosody in spoken English garden path sentences: ERP evidence for the boundary deletion hypothesis. *J. Cogn. Neurosci.* 23, 2731–2751.

Poplack, S., and Walker, D. (1986). "Going through (L) in canadian French," in *Diversity and Diachrony*, ed. D. Sankoff (Philadelphia, PA: John Benjamins), 137–198.

Pourquie, M., St-Denis, A., and Royle, P. (2016). Flexion verbale et structure argumentale dans les troubles primaires du langage. Available at: http://tinyurl.com/congres60.

Rossi, S., Gugler, M. F., Friederici, A. D., and Hahne, A. (2006). The impact of proficiency on syntactic second-language processing of German and Italian: Evidence from event-related potentials. *J. Cogn. Neurosci.* 18, 2030–2048. doi:10.1162/jocn.2006.18.12.2030.

Royle, P. (2011). On the existence of C/Ø alternations in French adjectives: Theoretical and empirical questions. in *Proceedings of the 17th ICPhS* (Honk Kong), 1730–1733. Available at: https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2011/OnlineProceedings/RegularSession/Royle/Royle.pdf [Accessed October 26, 2016].

Royle, P., Courteau, É., and Steinhauer, K. (in preparation). ERPs for the study linguistic morphology. in press.

Royle, P., Drury, J. E., and Steinhauer, K. (2013). ERPs and task effects in the auditory processing of gender agreement and semantics in French. *Ment. Lex.* 8, 216–244. doi:10.1075/ml.8.2.05roy.

Rugg, M. D. (1995). *ERP studies of memory.* Oxford, UK: Oxford University Press.

Sassenhagen, J., Schlesewsky, M., and Bornkessel-Schlesewsky, I. (2014). The P600-as-P3 hypothesis revisited: Single-trial analyses reveal that the late EEG positivity following linguistically deviant material is reaction time aligned. *Brain Lang.* 137, 29–39. doi:10.1016/j.bandl.2014.07.010.

Selkirk, E.O. 1981. *English Compounding and the Theory of Word-structure,* in: M. Moortgat, H. Van der Hulst & T. Hoestra (eds.) The Scope of Lexical Rules, Foris, Dordrecht.

Steinhauer, K. (2003). Electrophysiological correlates of prosody and punctuation. *Brain Lang.* 86, 142–164. doi:10.1016/S0093-934X(02)00542-4.

Steinhauer, K., Alter, K., and Friederici, A. D. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nat. Neurosci.* 2, 191. Available at: https://www.nature.com/articles/nn0299_191.

Steinhauer, K., and Connolly, J. F. (2008). "Event-related potentials in the study of language," in *Handbook of the neuroscience of language*, ed. B. Stemmer (New York, NY: Elsevier Ltd.), 91–104.

Steinhauer, K., and Drury, J. E. (2012). On the early left-anterior negativity (ELAN) in syntax studies. *Brain Lang.* 120, 135–162. Available at:

http://www.sciencedirect.com/science/article/pii/S0093934X11001246 [Accessed February 13, 2017].

Steinhauer, K., Drury, J. E., Portner, P., Walenski, M., and Ullman, M. T. (2010). Syntax, concepts, and logic in the temporal dynamics of language comprehension: Evidence from event-related potentials. *Neuropsychologia* 48, 1525–1542. doi:10.1016/j.neuropsychologia.2010.01.013.

Steinhauer, K., and Friederici, A. D. (2001). Prosodic boundaries, comma rules, and brain responses: The closure positive shift in ERPs as a universal marker for prosodic phrasing in listeners and readers. *J. Psycholinguist. Res.* 30, 267–295. Available at: https://link.springer.com/article/10.1023%2FA%3A1010443001646.

Steinhauer, K., Royle, P., Drury, J. E., and Fromont, L. A. (2017). The priming of priming: Evidence that the N400 reflects context-dependent post-retrieval word integration in working memory. *Neurosci. Lett.* 651, 192–197. doi:10.1016/j.neulet.2017.05.007.

Stowe, L. A., Kaan, E., Sabourin, L., & Taylor, R. C. (2018). The sentence wrap-up dogma. *Cognition*, *176*, 232-247. doi:10.1016/j.cognition.2018.03.011

Tanner, D. (2015). On the left anterior negativity (LAN) in electrophysiological studies of morphosyntactic agreement: A Commentary on “Grammatical agreement processing in reading: ERP findings and future directions” by Molinaro et al., 2014. *Cortex* 66, 149–155. doi:10.1016/j.cortex.2014.04.007.

Tanner, D., and Van Hell, J. G. (2014). ERPs reveal individual differences in morphosyntactic processing. *Neuropsychologia* 56, 289–301. doi:10.1016/j.neuropsychologia.2014.02.002.

Tokowicz, N., and MacWhinney, B. (2005). Implicit and explicit measures of sensitivity to violations in second language grammar: An event-related potential investigation. *Stud. Second Lang. Acquis.* 27, 173–204. doi:10.1017/S0272263105050102.

Van Berkum, J. J., Brown, C. M., and Hagoort, P. (1999). Early referential context effects in sentence processing: Evidence from event-related brain potentials. *J. Mem. Lang.* 41, 147–182. doi:10.1006/jmla.1999.2641.

van de Meerendonk, N., Kolk, H. H., Chwilla, D. J., and Vissers, C. T. W. M. (2009). Monitoring in language perception. *Lang. Linguist. Compass* 3, 1211–1224. doi:10.1111/j.1749-818x.2009.00163.

Willems, R. M., Özyürek, A., and Hagoort, P. (2008). Seeing and hearing meaning: ERP and fMRI evidence of word versus picture integration into a sentence context. *J. Cogn. Neurosci.* 20, 1235–1249. doi:10.1162/jocn.2008.20085

Ye, Z., Luo, Y. J., Friederici, A. D., & Zhou, X. (2006). Semantic and syntactic processing in Chinese sentence comprehension: Evidence from event-related potentials. *Brain research*, *1071*(1), 186-196. doi: 10.1016/j.brainres.2005.11.085

4.7. Supplementary materials manuscript 2

4.7.1. Sentence structures

180 French verbs acquired before the age of 8 years were selected from the Manulex database (Lété, Sprenger-Charolles, & Colé, 2004). Both Manulex (Lété, Sprenger-Charolles, & Colé, 2004) and Lexique (New, Pallier, Ferrand, & Matos, 2001) were consulted to provide oral language frequency norms for selected items, since the experiment will ultimately be used with children. Of the 180 verbs selected, one third were intransitive and were followed by a prepositional phrase (PP) (e.g., *Le sportif, il court dans le parc de la ville*, ‘The athlete, he runs in the city park’), one third were transitive (e.g., *La fille, elle aime la mousse au chocolat*, ‘The girl, she likes chocolate mousse’) and one third were ditransitive (e.g., *Le dompteur, il lance une balle à l'otarie*, ‘The tamer, he throws a ball to the see lion’). Transitive and intransitive verbs were followed by an NP or a PP (or both). Note that to maintain equal sentence length across verb argument-structure types, certain transitive verbs were followed only by an NP. These account for 19% (15/77) of transitive verbs or 8% (15/180) of the verbs used. Verbs were inserted into sentences containing singular or plural third person subject pronouns (*he/she/they*), and a sentence continuation phrase consisting of a direct object NP, or a prepositional phrase (e.g., … *in the public pool*), to postpone wrap-up effects in the ERPs (Hagoort, 2003). Context phrases (neutral – e.g., *Each week* – and subject NP – e.g., *The star*) were also created. We ensured that there were no additional instances of liaison in sentences or carrying phrases, such that only sentences in the liaison verb sub-condition contained them. We also ensured that all nouns, adverbs, prepositions and adjectives included in sentences and carrying phrases were age appropriate for children, as per Manulex (Lété et al., 2004). Further, subject grammatical gender

(feminine or masculine), as well as syllable length of context phrases and of full sentences were balanced across the three verb categories.

4.7.2. Stimulus recording

Auditory material recording, normalizing, and splicing was performed by trained research assistants with a background in speech editing. Auditory stimuli were recorded in a sound-shielded audiology booth using a Sony DAT recorder (PCM-M1 recorder, 1997). All sentences were spoken by a native French Canadian actor who was trained to pronounce words with clear but natural articulation, while maintaining constant and natural prosody and intonation. Sentences were spoken with natural within- and between-word co-articulation, however this was avoided at splicing points (i.e., between the last word in context phrases and the first word in the following sentences). The actor was also instructed to maintain a constant vocal intensity, intonation, and speech rate, throughout the recording. Voice-volume monitoring (+ or – 5 dB) was performed during recording. All conditions within a given block of stimuli were recorded together. All of this allowed us to create very natural-sounding auditory stimuli, while reducing post-recording manipulations.

4.7.3. Sound processing

Before cross-splicing took place, context phrase and sentence files were processed and normalised in five steps using Audacity® (Audacity Team, 2018) and Praat (Boersma & Weenik, 2018) software: (1) we reduced noise, such as microphone feedback and other extraneous sounds that were present in the recordings, by applying Audacity's built-in Noise Reduction function to all files. During this process, care was taken to use the lowest level of manipulations possible in an effort to reduce distortion to the resulting audio files. For example, the sensitivity (which controls how much of the audio is considered noise) and frequency smoothing (which spreads the noise reduction process to the specified number of neighbouring frequency bands) options were always maintained at lowest possible levels; (2) as noise reduction created new instances of 'silence' at the start and end of audio files, such silences were trimmed using the Audacity Trim Silence plug-in (by Daulton, 2011). A threshold of -35 dB was selected, as this level was deemed most appropriate in removing silent portions, while sparing phonemes with lower amplitudes such as voiceless stops, during tests performed on a subset of files; (3) using a Praat script (created in-house by the second author), sentence onsets and offsets were then trimmed at zero-crossings –the point where the waveform crosses the zero-level axis–, in order to avoid discontinuities in the sound wave which can be perceived as clicks or pops; (4) finally, we listened to each file to assess the naturalness of speech rate. In cases where speech rate was perceived to be too fast or slow in comparison to the majority of audio files, or where there was a discontinuity between context phrases and associated sentences, speech rate was adjusted by using Audacity's built-in *Change Tempo* function. The lowest level of manipulation necessary to create natural sounding and consistent speech was used (maximum 15% slower or faster); (5) lastly, we added silent portions to the onsets and offsets of all audio files, such that all final audio files (once context phrases and sentences were cross-spliced) contained 0.035 seconds of onset silence, 0.05 seconds of offset silence, and 0.1 seconds of silence between the context phrase and the rest of the sentence (these values were perceived to be most natural-sounding when tests were performed on a subset of audio files). Audacity's *Trim/Extend* plugin (Daulton, 2011) was used to process all files in this manner. Once context phrase and sentence processing was complete, they were cross-spliced to create audio files for different experimental conditions, as outlined in the next sections. Cross-splicing was carried out in Praat by means of a script (created in-house by the present author) which concatenated all context-phrase and sentence files based on conditions outlined above. Lastly, the amplitude of the resulting audio files was normalised to a level of 70 dB SPL using Praat.

4.7.4. Developing stimuli for the semantic condition

Items for the matched condition included 60 different verbs and were created in three steps. First, a native French Canadian actor spoke each congruent item, consisting of a context plus sentence combination. We then cross-spliced the neutral and subject NP contexts, as depicted in Table 1 of the article, examples (1a) - (1b), and (1c) - (1d).

Each item pair was associated with one image depicting the action described in the audio files. Items for the mismatched condition were created by cross-splicing the contexts from one item pair with the sentences from another, as depicted in examples (1c) - (1d), resulting in 120 mismatched item pairs. In the end, 60 item quadruplets were created for a total of 240 items (120 matched and 120 mismatched). In order to create semantic mismatches, each item quadruplet was associated to the same image, such that both matched and mismatched item versions were displayed with the same image. Table 1 provides an example of one item quadruplet, showing example auditory stimuli for each subcondition.

4.7.5. Developing stimuli for the number mismatch condition

First, we recorded a subset of congruent items including 120 different verbs, namely a singular subject NP context and singular sentence combination (as in Table 2 of our article, item 1b), a neutral context and singular sentence combination (as Table 2 of our article, item 1a), and a plural subject NP context and plural sentence combination (as in Table 2 of our article, item 2b). This resulted in 120 sentence triplets (120 with singular subject NP contexts and singular sentences, 120 with plural subject NP contexts and plural sentences, and 120 with neutral contexts and singular sentences), for a total of 360 spoken items. We then cross-spliced as in the example provided in Table 2: each singular sentence was spliced with both a singular subject NP and neutral context (as in 1a and 1b), and each plural sentence was spliced with a plural subject NP context and the same neutral context that was used for the singular (as in 2a and 2b). This created item quadruplets, whereby the two singular versions differed in context only, the two plural versions also differed in context only, and the same neutral context was paired with both singular and plural sentences. This resulted in 120 item quadruplets, or 480 items. Each singular and plural item pair was associated with a corresponding image (i.e., singular items were paired with an image depicting one agent, and plural items were paired with an image depicting two agents). Number mismatches were created by presenting the same grammatical audio files, but swapping the images, such that audio files describing singular subjects were presented with images depicting plural subjects (Table 2 of our article, items 1c-d and 2c-d), and vice versa. In total, 960 audio file and image combinations were created (480 congruent and 480 incongruent).

4.7.6. List creation

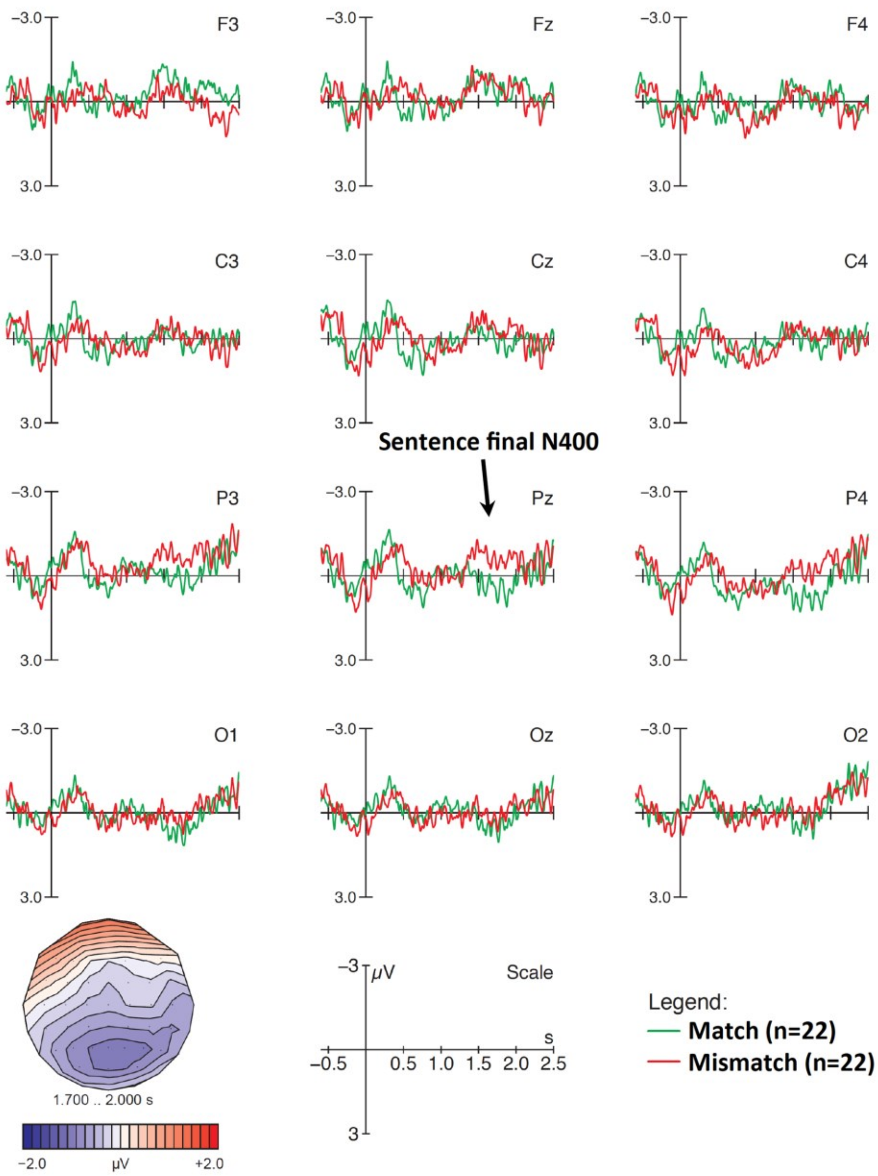
Four different presentation orders (lists) were created. In each list, stimuli were evenly distributed across 15 blocks of 20 items each. The following constraints were met: (1) each block contained two items in each of the eight morphosyntactic sub-conditions as well as one sentence in each of the four semantic sub-conditions; (2) there was no consecutive repetition of the same sub-condition; (3) match and mismatch conditions were evenly distributed across each block; (4) in order to minimize strategic processing effects, pseudorandomization within blocks also prevented (a) consecutive presentation of items with the same agent, (b) consecutive presentation of more than three items from congruent or incongruent conditions, (c) consecutive presentation of more than three singular or plural items, (d) clusters of particularly long or short sentences. In order to further rule out any sequence effects, four additional mirror versions of each list were created by reversing both the block order and the sentence order within each block. Thus, a total of eight experimental lists were created and evenly assigned across male and female participants.

4.7.7. Sentence wrap-up effects for number mismatch conditions

As in the semantic mismatch condition, number mismatches also elicited posterior negativities at sentence-final positions that are compatible with a wrap-up N400 effect (see Supplementary Figures 1 and 2). A global ANOVA on the 1700–2000 ms time-window comparing match and mismatch verbs (both CONS and LIAIS) yielded significant CONDITION× ANTERIORITY ($F(2, 42) = 4.19$, $p = 0.04$), and CONDITION× LATERALITY ($F(1, 21) = 6.36$, $p = 0.02$) interactions in the lateral electrodes. These interactions reflect the negativity's posterior and medial distribution. While no significant interaction with CONTEXT was found, separate plots for the two contexts suggested that the wrap-up effect was more prominent for NP contexts. In fact, separate analyses revealed no clear wrap-up N400 effect for neutral contexts. In contrast, a comparison of match and mismatch verbs within NP contexts again reveals significant interactions of CONDITION× ANTERIORITY ($F(2, 42) = 5.56$, $p = 0.02$), and CONDITION× LATERALITY ($F(1, 21) = 4.70$, $p = 0.04$), in lateral electrodes. These reflect the fact that a main effect of CONDITION is observed in posterior electrodes ($F(1, 21) = 4.60$, $p = 0.04$).

Supplementary Figure 1

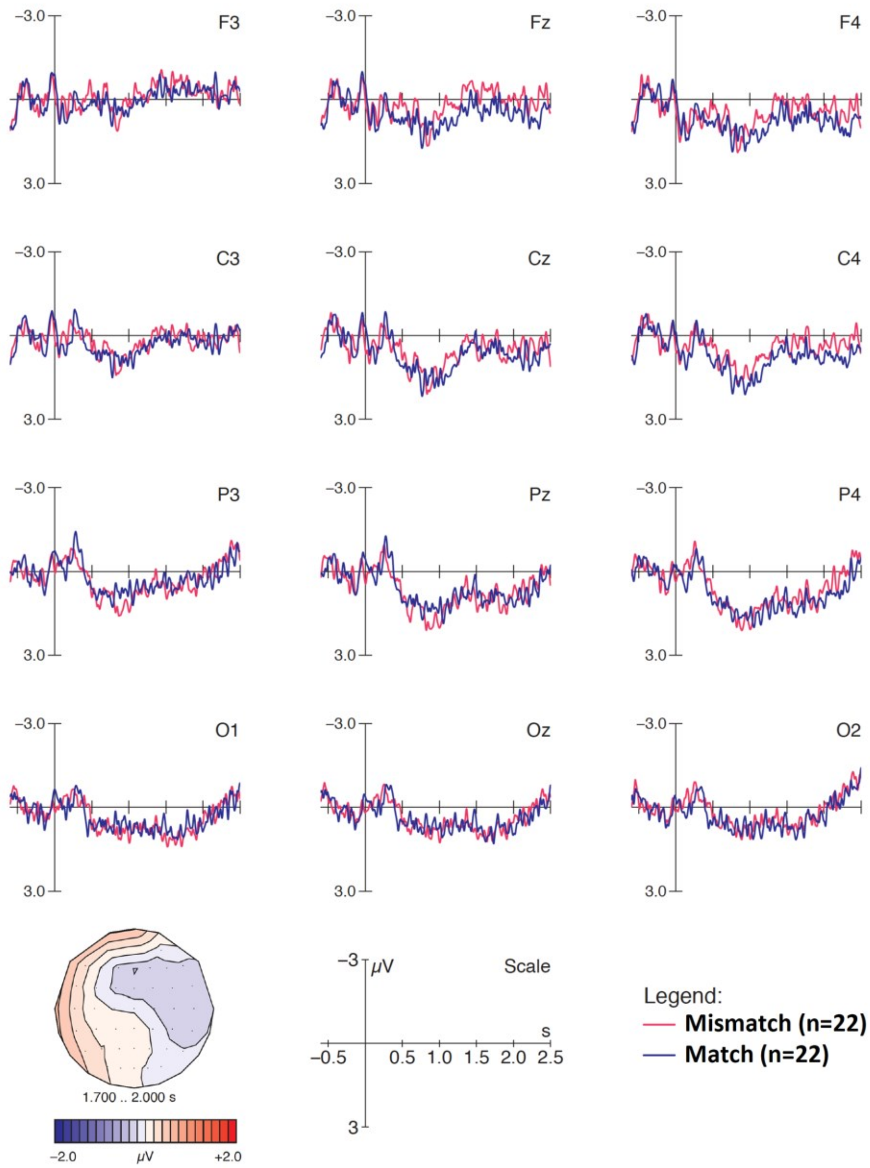
Sentence wrap up effects in number mismatch conditions for NP contexts



Note. Late effects for cross-modal number mismatches in NP contexts measured from the onset of the verb. Displayed are grand-average ERPs at midline and medial electrodes for all participants, time-locked to the onset of the verb (vertical bar) with a baseline of -600 to 0 ms. Mismatching NP context conditions (red) elicited a late right-lateralized posterior negativity in the 1700–2000 time-window relative to matching NP contexts (green).

Supplementary Figure 2

Sentence wrap up effects in number mismatch conditions for neutral contexts.



Note. Late effects for cross-modal number mismatches in neutral contexts measured from the onset of the verb. Displayed are grand-average ERPs at midline and medial electrodes for all participants, time-locked to the onset of the verb (vertical bar) with a baseline of -600 to 0 ms. Mismatching neutral context conditions (magenta) elicited no negativity in the 1700–2000 time-window relative to matching neutral contexts (blue).

4.7.8. References

Audacity Team (2017). Audacity(R): Free Audio Editor and Recorder [Computer application]. Version 2.2.1 retrieved December 20th 2017 from https://audacityteam.org/

Boersma, P., and W., David (2018). Praat: doing phonetics by computer [Computer program]. Version 6.0.40, retrieved September 2018 from http://www.praat.org/.

Lété, B., Sprenger-Charolles, L., and Colé, P. (2004). MANULEX: A grade-level lexical database from French elementary-school readers. *Behavior Research Methods, Instruments, & Computers*, *36*, 156–166. doi: 10.3758/bf03195560

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. doi:10.1162/089892903322370807

5. Bridge 2

With the second manuscript, we demonstrated that our novel experiment using only grammatical sentences elicited classic ERP components found in experiments with ungrammatical sentences in French-speaking adults. Furthermore, our results indicated that different patterns are to be expected respectively in the lexico-semantic and the morphosyntactic conditions.

We will now turn to the investigation of morphosyntax and lexico-semantic processing at the neurocognitive level in (pre-)teenagers with and without DLD. We will investigate language processing with our novel ERP experiment in the context of the predictions based on two neurocognitive models of DLD, the Procedural Deficit Hypothesis (PDH) and the Generalised Slowing Hypothesis (GSH).

6. Manuscript 3

Testing neurocognitive models of language processing in developmental language disorder

Émilie Courteau[1,2], Phaedra Royle [1,2], Karsten Steinhauer[2,3]

[1]School of Speech Language Pathology and Audiology, Faculty of Medicine, University of Montreal, Montréal, Québec, Canada

[2]Centre for Research on Brain, Language and Music (CRBLM), Montréal, Québec, Canada

[3]School of Communication Sciences and Disorders, Faculty of Medicine, McGill University, Montréal, Québec, Canada

This manuscript in under revision for the Journal of Neurodevelopmental Disorders

Abstract

Background: A recent label change from Specific Language Impairment (SLI) to Developmental Language Disorder (DLD) has been proposed for neurodevelopmental language impairments. Grammatical impairments have been identified as the hallmark of DLD. The procedural deficit hypothesis proposes that these impairments are related to a procedural memory deficit, with a preserved declarative memory for lexico-semantic abilities. The new DLD definition states that all linguistic domains can be impaired. This reopened the door for accounts that extend to all areas of language, such as the generalized slowing hypothesis, which posits that a processing speed deficit underlies DLD. Our study contrasts these two views using an event-related potential (ERP) experiment of sentence processing in French-speaking (pre-)teenagers with and without DLD.

Methods: Two groups of (pre-)teens, with DLD or typical language development (TL), listened to grammatical sentences describing depicted images while their EEG was recorded. Lexico-semantic mismatches were created by presenting a verb that did not match the depicted action. Number agreement mismatches were created by varying the number of visually presented subjects and morphosyntactic number cues in the auditory stimuli, either in the noun phrase at sentence onset, or on the verb downstream in the sentence. Subject-verb number agreement was marked on plural verb forms through regular liaison or with irregular verb-final consonants.
Results: Accuracy of acceptability judgements revealed that while the TL group clearly outperformed the DLD group in morphosyntactic conditions, both groups performed similarly in the lexico-semantic one. ERPs revealed similar patterns between groups in the lexico-semantic (indexed by N400s) and number agreement at sentence onset conditions (indexed by P600s), but with delays for the DLD group. Both groups elicited an N400 in response to irregular plural verb

agreement mismatches, while different patterns were found on regular liaison plural mismatches. No biphasic ERP pattern reminiscent of adult-like processing was found in either group.

Conclusion: Although some processing delays were found in the DLD group, differential patterns on most conditions better fit the procedural deficit hypothesis. Our study suggests that, in contrast with morphosyntax, lexico-semantics is a relative strength in teenagers with DLD when processing linguistic information at the sentence level.

Keywords: Developmental language disorder, Specific language impairment, Adolescents, Sentence processing, Event-related brain potentials, N400, P600, French language, Neurocognition

6.1. Introduction

Many researchers have proposed that children with language impairments have limited knowledge of grammatical rules or constraints, resulting in production and comprehension problems (for reviews see Pourquié et al., in revision, and Jakubowicz, 2003). These accounts support the idea that only one or a subset of language subdomain(s) need to be impaired for a children to be diagnosed as with a specific language impairment (SLI). An example of this approach is the SLI group studied by van der Lely (1998), which was presented as having only a deficit in the grammatical domain and was hence labelled G-SLI (grammatical-SLI). As noted by Ullman and Pierpont (2005), some limitations of these accounts lie in the fact that they explain only the behavioural performance of children with SLI and not their underlying neurocognitive processes, and ignore the nonlinguistic deficits that often co-occur with SLI. The procedural deficit hypothesis (PDH, (Ullman & Pierpont, 2005; Ullman et al., 2020)) overcome these limitations by taking into consideration both the evidence of grammatical deficits of SLI in presence of lexico-semantics strengths, and deficits in other cognitive domains such as motor and non-verbal memory.

Even if the PDH explains a broader range of SLI's behavior, it maintains that some language abilities are spared by the impairment, as for instance lexico-semantics. However, from a clinical point of view, it is known that most children with language impairment have difficulties in all linguistic domains. This reality aligns with the recent label change from SLI to developmental language disorder (DLD, Bishop et al., 2017) proposed by a consortium of clinicians and researchers. Functional impairments, language deficits that impact all educational progress, and co-morbid diagnoses are now accepted and expected in individuals with DLD. This definition reopened the door for accounts that include deficit in all linguistic domains, such as the General Slowing Hypothesis (GSH) proposed by Kail in 1994. Kail suggested that DLD[1] is the result of a common slowing coefficient that modulates all cognitive functions in children with DLD, pointing to a general processing deficit.

While several studies have compared lexical-semantic and morphosyntactic skills in children with DLD, few simultaneously examine the timing underlying processing of these two domains. We propose a study that contrasts the perspectives of the PDH and the GSH in a bimodal audio-visual event-related potential (ERP) experiment that assesses lexical-semantic and morphosyntactic processing at a behavioural and neurocognitive level. French-speaking (pre-)teenagers with and without DLD participated in this study. French provides the opportunity to study these dissociations as this language has a rich morphosyntactic system. Indeed, grammatical markers such as number agreement are acquired and consolidated from early childhood to school-age years even in children with typical language (TL).

6.1.1. Two accounts for developmental language disorder

In the DSM-V (American Psychiatric Association, 2013), the diagnostic criteria for children with language disorder include early onset of symptoms and persistent difficulties in the acquisition of language due to deficits in comprehension or production, characterised by a reduced vocabulary, limited sentence structures and discourse impairments. The new DLD label proposed by Bishop et al. (2017) aligns with the DSM-V definition and provides additional guidelines for clinicians. For example, the first step toward a DLD diagnostic must be the presence of a functional impairment and comorbid conditions must not prevent the diagnosis.

[1] Note that when we review studies that use the previous common label Specific Language Impairment (SLI), we will use the label DLD for the sake of simplicity.

The PDH posits that abnormalities in the DLD brain structures supporting procedural memory can explain many of their behavioural characteristics (Ullman et al., 2020; Ullman & Pierpont, 2005). Among others, the procedural memory system plays a role in morphosyntactic processing. Within morphosyntax, at the word-form level, the procedural system supports ruled-govern regular complex word processing such as in English third person verb number agreement (e.g., she sings) and, in French, plural verb number agreement liaison instantiated by liaison of the plural pronoun followed by a verb with a vowel onset (e.g., elle achète [ɛlaʃɛt] 'she buys' vs. elles achètent [ɛlzaʃɛt] 'they buy'). Unpredictable word form as in English she sang, or in French ils pondent [ilpɔ̃d] ('they lay eggs') are argued to be stored in the lexicon, which is supported by the declarative memory system. At the sentence-level, the procedural system underlies agreement between sentence constituents such as number or tense (Steinhauer & Ullman, 2002). In children with TL, ruled-governed morphosyntactic information either at the word-form or sentence level is assumed to first be supported by the declarative system. It is gradually taken over by the procedural system, which lead to automatization (Ullman, 2020). In the case of DLD, the procedural memory deficit would mean that rule-governed morphosyntactic information will not be processed by this system; instead, this type of information will be stored and processed by the declarative memory system as a compensatory strategy (Ullman et al., 2020; Ullman & Pierpont, 2005).

Note that within lexico-semantics, the procedural memory is expected to underpin lexical retrieval (Ullman and Pierpont, 2005). It is thus expected that tasks involving lexical retrieval are likely to be more affected in DLD than receptive tasks such as word comprehension or word recognition. In addition, if lexical retrieval tasks do not require rapid responses, the PDH predicts that improved and possibly normal performance should be observed in participants with DLD (*ibid*, p. 418).

The PDH is supported by evidence at the neurocognitive level, which has shown atypicality of the procedural system in brains of children with DLD, and at the behavioural level, with evidence for deficits in (morpho-)syntax and multiple non-linguistic domains supported by the procedural system (for a review see Ullman et al. 2020, p. 15). The portrait is less clear regarding the evidence supporting the integrity of the declarative system as illustrated by strength in lexico-semantics. Similar neurocognitive processes underlying lexico-semantics have been observed between children with and without DLD but sometimes with delays (e.g. see next section for a review). Results from behavioural tasks assessing stored conceptual and lexical knowledge sometimes show deficits in children with DLD, as in acceptability judgments of sentences containing lexico-semantic anomalies on nouns and verbs (Pawlowska et al., 2014; Haebig et al., 2017). However, the PDH predicts similar performance between children with and without DLD for anomalies of this kind (Ullman & Pierpont 2005, p. 415).

The GSH (Kail, 1994) proposes that children with DLD use the same processes as TL ones to execute a task.; however, the time to execute each process will be multiplied by a common coefficient resulting in a generalized slowing effect. A slower processing speed will result in processing limitations because, as the information is not processed fast enough, it will be vulnerable to degradation or interference from other incoming information (Kail & Salthouse, 1994). Originally, Kail (1994) proposed that DLD's slower processing speed would affect linguistic as well as non-linguistic domains in a similar way. However, Leonard et al. (2007) showed with multiple tasks administered to 14-year-olds with and without DLD, using confirmatory factor analyses, that linguistic processing speed can be differentiated from non-linguistic ones, and therefore a same slowing factor cannot apply to all cognitive domains. Their results indicated that processing speed should be considered as a dimension distinct from verbal and nonverbal working memory. Nevertheless, these authors concluded that processing limitations can mitigate children's language difficulties and, perhaps, be a primary cause. Regarding linguistic tasks, several studies reported slower response times for children with DLD when compared to their peers with TL in many linguistic domains (e.g., in English: Miller et al. 2006; in French: Quémart & Maillart 2016, Royle et al. 2002). A limitation of these studies, as noted by Kail (1994), is that they reported a delay based on group averages, so it is impossible to know whether a delay is present in all children with DLD. Windsor and Hwang (1999) addressed this issue by analyzing response times at the individual level on tasks that assessed linguistic and non-linguistic skills and showed that delays didn't affect a subgroup of DLD participants: 7 participants of the 23 with DLD showed faster response times than the control-group mean. In sum, evidence shows that many, but not all, children with DLD have slower processing speeds. It is noteworthy that the studies examining the GSH did not discuss Kail's (1994) proposal that similar processes should be used to complete a task in children with and without DLD. For instance, Miller et al. (2006) reported slower response times on all their tasks in teenagers with DLD when compared to the control one but no differences between groups were found on accuracy in 3 of 8 tasks. Indeed, both groups performed similarly on picture naming and lexical matching tasks as well as on visual search task, but not on grammatical and phonological tasks. The authors did not discuss how these differences in task performance across domains fit with the GSH framework, but these results suggest differences between the groups in the processes underlying the grammatical and phonological tasks.

Overall, the two accounts differ in how they view language processing by children with DLD. The PDH posits that DLD preserves the declarative system underlying many aspects of lexical-semantics and will take over the processing of linguistic domains that are usually subserved by the procedural system, such as morphosyntax. When processing morphosyntax, it is thus expected that children with and without DLD won't use the same processes. The GSH proposes that children with DLD will use the same processes as their peers with TL to complete a linguistic task, but with a limited processing speed. Interestingly, on the one hand, the PDH does not appear to have a stake on the significance of the observed delays in DLD response times, i.e., whether this indicates a deficit. On the other hand, studies supporting the GSH interpret only the delays in DLD task performance and do not appear to integrate into their interpretations data pointing to impaired processes underlying delayed responses.

The ERP technique is a suitable way to test these two accounts' views on DLD. Distinct ERP profiles can reflect lexical-semantic and morphosyntactic processing differently, and this technique offers an excellent temporal resolution of less than one millisecond which also allows for fine-grained analyses of the time course of these two types of processing.

6.1.2. Tracking lexico-semantic and morphosyntactic development with ERPs

Using electrodes caps, the electroencephalogram (EEG) records neural responses related to linguistic events in the form of event-related potentials (ERPs). ERPs, or ERP components, are described in terms of polarity, as the electrical potential can be negative or positive, timing, as these electrical potential flows over time, and in terms of topography, where they typically emerge over the scalp surface. Over the last 20 years, many ERP studies have evaluated lexico-semantic and morphosyntactic development in children. For the purpose of this article, we will concentrate our review on studies of older children and adolescents with and without DLD.

Where not specified, the language of study participants is English. For an in-depth review including an analysis of methodological flaws often observed in ERP studies of children, see Royle and Courteau (2014), and for a review of ERP and MEG studies focusing on morpho-syntax, see Royle and Steinhauer (in press).

Lexical-semantic processing is typically reflected by a centro-parietal negative brain wave between 300–500 ms after word onset, the N400 (Kutas and Hillyard, 1980). This component can be elicited by lexical-semantic expectancy violations (Kutas and Federmeier, 2011; Steinhauer and Connolly, 2008) in reading, auditory or bimodal auditory-visual modalities, where a mismatching image is presented concurrently with an auditory utterance, for instance Chaque semaine, elle !chante .. 'Each week, she !sings…' with an image of a [woman that SWIMS] (Courteau et al. 2019; see also Friedrich and Friederici 2004, and Courteau et al., 2013). By the end of their second year, children can relate and integrate aspects of lexico-semantic information as represented by the N400 during bimodal stimuli presentation of isolated words (for a review see Morgan et al., 2020). N400s similar to adults in terms of latency and distribution are observed in children from 7 years of age (Cummings et al., 2008). Whereas the absence of N400s in 19-month-old infants was related to low scores on word production tasks at 30 months and thus to a risk of developing DLD (Friedrich and Friederici, 2004, 2006), N400s in response to lexico-semantic anomalies are consistently observed in children and teens with DLD. Neville et al. (1993) found N400 onset delays as well as topographical differences in 9-year-old children with DLD, compared to typically-developing children, during sentences reading with semantically anomalous final nouns (e.g., Giraffes have long !scissors). However, the details of this delay remain unclear as the authors didn't provide statistics for all time-windows analysed. Delays in N400 onsets were also found by Pijnacker et al. (2017), in a group of Dutch children with DLD aged 5 when compared to children of the same age with TL. Children listened to simple sentences with final semantically incongruent nouns, while watching unrelated silent short video clips. The authors found that the N400 onset varied between 300–500 ms for the control group, and 500–800 ms for the group with DLD. The N400 was broadly distributed over the scalp for the DLD group while it was posterior in the typical children. Furthermore, they found that in the DLD group, smaller N400 amplitudes were associated with lower scores on tasks assessing grammar, vocabulary, language comprehension and nonverbal IQ, with small to moderate correlations. Using a bimodal picture–noun paradigm with semantically unrelated material to create mismatches, Kornilov et al. (2015) found similar N400 onsets (310–410 ms) in a group of Russian-speaking participants with DLD aged 10 years on average, and a control group of age-matched children, but with an earlier N400 offset for the DLD group. The authors didn't find significant correlations between scores on lexical, grammatical, or phonological tasks and the N400 amplitude. In a study of auditory-sentence processing with no visual support, Haebig et al. (2017) found no timing or duration differences, measured using sequential temporal analyses of 50 ms time-windows, in 16 year-olds with and without DLD on verb with lexico-semantic anomalies (e.g., '…the horse !sings'). It remains unclear if there were differences between groups on scalp distribution of effects. The authors mentioned a significant interaction including group and topographical factors but didn't report the decomposition of these effects. Overall, N400s reflecting lexico-semantic expectancy violations are found in participants with DLD, but occasionally with quantitative difference with their TL peers. The authors reported N400s with a delayed onset or with a reduced duration, and different scalp distributions, as one study reported a more focal distribution for participants with DLD.

In adults, studies have shown that morphosyntactic processing usually elicits first a negativity linked to automatic grammar processing (Steinhauer and Connolly, 2008). This negativity is known to take different forms depending on the type of morphosyntactic error and written or auditory stimulus modality. Typically, morphosyntactic processing is known to be reflected between 300 and 500 ms by a negativity focused in the left anterior region of the scalp, the left anterior negativity (LAN; see Molinaro et al., 2011, for a review and Caffarra et al., 2019, for a recent discussion on the controversy regarding the reliability of the LAN). LANs have been observed mainly in the written modality. In the auditory modality, bilaterally-distributed anterior negativities (AN) with a longer duration, up until 1200 ms post stimuli, had been observed in response to subject-verb number agreement processing, as in French during the presentation of ungrammatical auditory sentences (Isel and Kail, 2018) or of mismatches between grammatical auditory sentences and a picture depicting the wrong subject number (Courteau et al., 2019). In this latter study, audio-visual mismatches on regular and irregular subject-verb number agreement elicited sustained anterior negativities while regular determiner-noun number agreement in noun phrases (NP)[2] elicited N400s. The LAN, AN and N400 for morphosyntactic processing are usually accompanied by a P600, a positive wave typically 600–1000 ms after stimuli onset (Steinhauer & Connolly, 2008). The P600 has been associated with late, and probably controlled, sentence reanalyses and sentence categorisation as incorrect (Bornkessel-Schlesewsky and Schlesewsky, 2019; Royle et al., 2013).

Studies on morphosyntactic processing in children first reveal the absence of a clear trend in the age at which children display the same ERP components as adults. Consider, for example,

[2] We are using NP as a shorthand for a complex noun phrase [DP Det [AP Adj [NP Noun]]].

the most studied morphosyntactic marker in child ERP studies: third person number agreement in English and commission errors (e.g., the boys often *cooks) or omission (e.g., the boy often *cook). In Dube et al. (2016), when adults heard these errors in a visual-auditory paradigm, an anterior negativity followed by a P600 was elicited. In TL children aged 9 to 11, commission errors elicited only a P600 while omission errors elicited no effect (Dube et al., 2018). However, in a study with a similar experimental design, Purdy et al. (2014) found in 7 to 11;5-year-old children with TL an anterior negativity and a P600 in response to auditory commission errors, a pattern similar to Dube et al.'s (2016) adult group. Using only auditory sentences, Weber-Fox et al. (2010) first found an anterior negativity in response to commission errors in both control and DLD groups aged 14;3 to 18;1-years-old. However, based on a visual inspection, this bilaterally distributed negativity in both anterior and posterior channels from 350–550 ms could be interpreted as a broadly distributed small N400 that did not reach significance in the midline channels. Following this negativity, a P600 was found in the control group only. Also using only auditory sentences, Haebig et al. (2017) tested three groups of teenagers aged 16-years-old on commission and omission errors. Both types of errors elicited only P600s in their TL group, and no effect for their DLD group. However, their group of adolescents with recovered DLD, who overcame their language impairment from childhood to adolescence, showed an N400 in response to commission errors but a P600 in response to omission errors. This suggests that TL teens processed errors as ungrammatical (as indexed by a P600 for sentence reanalyses), that participants with DLD did not process these errors, and that those with recovered DLD processed commissions as lexical errors but omissions as grammatical ones.

Based on these studies, we note that age doesn't seem *per se* to impact ERP responses in these (pre-)teen groups. However, we can see that depending on the study and the experimental group, third person number agreement errors elicited either no effect, an N400 only, a P600 only, or an adult-like biphasic pattern, i.e., a negativity (AN) followed by a positivity (P600). Furthermore, no studies found this biphasic pattern in their DLD group. An examination of ERP studies that have assessed morphosyntactic processing in children or adolescents using different structures or in other languages also suggests that the biphasic pattern is absent in children or teens with DLD (see e.g., Cantiani et al., 2015). The ERP components elicited during morphosyntactic processing suggest at which stage of morphosyntactic development the participants in these studies were. Indeed, Steinhauer et al. (2009, p. 31) propose a learning trajectory for morphosyntactic ERPs based on studies of second-language learners. They suggest that ERPs can reflect different stages of morphosyntactic proficiency, from no perception of morphosyntactic errors, as evidenced by no ERP differences between correct and incorrect conditions, to native-like ERPs represented by the biphasic LAN-P600 pattern. See Table 1 for a review of ERP studies on morphosyntactic processing in children with and without DLD focusing on stages children go through during their mastery of morphosyntactic processing, and adapted from the learning path from novice to native-like of second-language learners proposed by Steinhauer et al. (2009).

Table 1

Developmental level of morphosyntactic processing as assessed through event-related potentials

Development level	Cognitive process	ERP pattern	References: Population/ age[years; months]/ agreement/ language
1-Immature	Indifferent perception	No effect	[a] Dube et al. (2018): TL / 9-11 / verb, 3rd person -s., omission / bimodal / English [b] Haebig et al. (2017): DLD / M = 16;5 / verb, 3rd person -s., commission + omission / auditory / English
2-Very low	Difficulties during lexical access and integration; compensatory processing strategy, likely relying on semantic plausibility and pragmatics	N400 or broadly distributed negativities	[c] Clahsen et al. (2007): TL / 6-7 / noun, regular, plural / auditory / German [a] Dube et al. (2018): TL / 9-11 / verb, 3rd person -s., commission / bimodal / English [b] Haebig et al. (2017): recovered DLD / M = 15;9 / verb, 3rd person -s., commission / auditory / English [e] Weber-fox et al. (2010): DLD / M = 15;9 / verb, 3rd person -s., commission / auditory / English[1] [f] Tippman et al. (2018): TL / 6-7 / verb tense, commission / written / German
3-Low to Intermediate	Beginning of grammaticalization	Small, frontal or delayed P600	[g] Cantiani et al. (2015): DLD / 8-13 / verb, regular subject-verb number / auditory / Italian[2] [h] Courteau et al. (2013); 2015: TL / 4;6-8;9 / determiner, gender / bimodal / French [i] Purdy et al. (2014): recovered DLD / 7-11;5 / verb, long-distance 3rd person -s., commission / bimodal / English [j] Schneider & Maguire (2019): TL / 8-9, 12-13/ verb, -ing commission + omission / auditory / English
4-Intermediate	Late controlled processing: repair approaches target mechanisms	Larger/ earlier P600	[g] Cantiani et al. (2015): TL/ 8-13/ verb, regular subject-verb number / auditory / Italian [b] Haebig et al. (2017): TL / M = 15;9 / verb, 3rd person -s., commission, omission / auditory / English [b] Haebig et al. (2017): recovered DLD / M = 15;9 / verb, 3rd person -s, omission / auditory / English [i] Purdy et al. (2014): TL/ 7-11;5/ long distance 3rd person -s., commission / bimodal / English [f] Tippman et al., (2018): TL / 8-9/ verb tense/ written / German
5-Near Expert	Near expert-like processing; early automatic processing + late controlled processing	Delayed or broadly distributed negativity + P600	[c] Clahsen et al. (2007): TL, adults / 8-9/ noun, regular, plural / auditory / German [d] Courteau et al. (2018): TL / 5-9 / adjective, gender / bimodal / French
6-Expert	Target mechanisms: early automatic + late controlled processing	AN/N400 + P600	[h] Brucher et al. (2020): adults / determiner, gender / bimodal / French [g] Cantiani et al. (Cantiani et al., 2013): adult s/ regular subject-verb number/ auditory / Italian [c] Clahsen et al. (2007): TL, adults / 11-12/ noun, regular, plural / auditory / German [a] Dube et al. (2016): adults / verb, 3rd person -s, commission, omission / bimodal / English [i] Purdy et al. (2014): TL and recovered DLD 7-11;5 / verb, 3rd person -s., commission / bimodal / English [f] Regel et al., (2015): adults / verb, tense / written / German [d] Royle et al. (2013): adults / adjective, gender / bimodal / French [e] Weber-fox et al. (2010): TL / M = 15;9 / verb, 3rd person -s., commission / auditory / English [j] Schneider & Maguire (2019): adults / verb, -ing ., commission + omission / auditory / English (P600 only)

Note. Letters identify studies using the same experiment but with different populations. TL = participants with typical language; DLD = participants with developmental language disorder; recovered DLD = participants that overcome language impairments. [1]Statistical analyses indicated that the negativity was significant only in the medial electrodes, and we argue it could be interpreted as a small N400 instead of an anterior negativity as suggested by the authors. [2]Visual inspection of ERPs suggests a later onset of the DLD group's P600, but the authors didn't do analyses of onset timing.

At the first level of development, an ungrammatical morphosyntactic maker is not recognized as such and there are no differences in brain activity between incorrect and correct control conditions. At the next level, morphosyntactic and lexico-semantic anomalies elicit N400s. As detailed by Steinhauer et al. (2009), the morphosyntactic violation is not recognized as such yet and the anomaly is therefore perceived as a lexical error (e.g., incorrect number agreement on nouns by German-speaking children aged 6-7 years old, Tippmann et al., 2018). At the third stage, children start to process the structural nature of the problem and attempt to reanalyze or repair it, leading to P600s that are small, delayed, or sometimes frontally distributed. Processes underlying this positivity could be different than the ones associated with a mature P600 (Steinhauer et al., 2009). Cantiani et al. (2015) found that Italian-speaking children with and without DLD aged 8 to 13 years old elicited a P600 in response to auditory subject-verb agreement errors. The P600 in the DLD group appears to be delayed in comparison to the control group, however the authors didn't analysed it. Schneider and Maguire (2019) presented auditory sentences with tense agreement present participle commission and omission errors (e.g., 'the hose can spray/*spraying') and found a frontally distributed positivity, in TL children aged 8-9 and 12-13, in contrast to a P600 in their adult group. However, this study lacks statistical and visual support. In French-speaking children with TL aged 5 to 9, gender agreement errors on the determiner in auditory noun-phrases, presented in an audio-visual paradigm, elicited a delayed P600 in comparison to a biphasic AN-P600 pattern in adults (Courteau et al., 2013, 2015; Brucher et al., 2021). At the next level of mastery, children should elicit a P600 similar to adults in response to morphosyntactic violations, usually with a parietal maximum, pointing toward mature and controlled sentence reanalyze processes. For instance, Tippmann et al. (2018) found, in response to incorrect verb tense agreement, a P600 with a similar timing onset between a group of German-speaking children with TL aged 8 to 13 years and adults (Regel et al., 2015). In the last two developmental stages, children will display biphasic ERP patterns like adults. In the fifth one, negativities and positivities could have a later onset or with different scalp distributions than adults. Using a visuo-auditory paradigm, Courteau et al. (2018) presented auditory sentences with gender agreement errors on adjectives to French-speaking children aged 5 to 9 years old. While adults displayed a bilateral anterior negativity and a P600 (Royle et al., 2013), children with higher levels of proficiency as assessed through behavioural tasks exhibited a delayed N400 followed by a delayed small P600. At the final and sixth developmental level, children are expected to display the same ERP patterns as adults, which is typically a biphasic AN or N400 followed by a P600 in response to morphosyntactic agreement anomalies (see Table 1 for studies finding biphasic patterns in children).

When assessing children's morphosyntactic processing using ERPs, we see that children with TL go through several stages of development that differ qualitatively or quantitatively from adults. However, based on our review, adolescents with DLD don't reach adult-like levels of processing as usually signalled by a biphasic ERP pattern. Indeed, qualitative differences are still observed, pointing toward the use of different or delayed processes. As evidenced by our short review, this doesn't seem to apply to lexico-semantic processing, as similar ERP components with occasionally quantitative, but not qualitative, differences have been observed. Overall, the ERP literature suggests differences between children and adolescents with and without DLD based on the type of linguistic information processed. However, these differences, such as temporal delays, are not systematically analyzed or discussed by studies. We propose a study that will address these differences.

6.1.3. The present study

The purpose of this study was to determine, using an ERP experiment, whether sentence processing by French-speaking teenagers with DLD is best characterised by the PDH or the GSH predictions. The PDH predicts that processes underlying lexico-semantics will be similar between children with and without DLD whereas they will be different when processing morphosyntax. The GSH proposes that both groups will use similar processing approaches for both linguistic domains, but they will be characterised by limited processing speed for the DLD group, instantiated by longer response times and timing delays. We tested these predictions at the behavioural level, via acceptability judgments, and at the neurocognitive level, measured with ERPs.

Contrary to much of the reviewed literature, our study used only grammatical sentences. Lexical-semantic mismatches were created between visually presented actions and spoken verbs, while auditory-visual subject number mismatches (singular vs. plural) were created by varying the number of visually presented subjects and morphosyntactic number cues in the auditory stimuli. We expected that this more naturalistic experimental paradigm would yield results more readily generalizable to everyday language use than the artificial context of ungrammatical sentences (Kandylaki and Bornkessel-Schlesewsky, 2019). Indeed, the combination of pictures and speech is similar to other common activities, such as shared picture-book reading or watching movies or videos, where a picture is presented along with a narration. In these cases, people being read or spoken to might make have expectations about what they will hear, and notice any incongruencies, as participants were expected to do during this experiment.

The assessment of morphosyntactic processing with ERPs in children with and without DLD underlies the evaluation of its acquisition, or at least its automatization. As seen in Table 1, close to adult or adult-like morphosyntactic processing, as illustrated by a biphasic ERP pattern, was observed for number agreement on nouns in children with TL starting at 8 years-old (Clahsen et al., 2007). For subject-verb agreement, the youngest group to display adult-like ERP was aged 7;5 to 11 years old (Purdy et al., 2014). However, with similar anomalies and stimuli, Haebig et al. (2017) observed only a positivity, reflecting an intermediate development level, in a group of teenagers with TL aged in average 15;9 years-old. Given that the type of number agreement (i.e., omission vs. commission) and the structure (i.e., NPs or verbs) seems to influence morphosyntactic processing, we developed an experiment based on French that used both types of agreement and structures. We used number agreement on determiners in NPs, which is understood and produced at high levels of mastery in children ages 1;8 to 3 years old (Valois et al., 2009). We also used subject-verb number agreement instantiated through regular liaison, which is understood by 3 year-old children in France, although it is not productively used even in adults (Legendre et al., 2014), and with irregular agreement, which is mastered in production around age 8 or later (Courteau et al., in revision; Franck et al., 2004).

Based on previous studies (see Table 1 and review in Royle and Courteau, 2014), adult-like ERPs can be observed starting at 8 years old, so this experiment targeted older children and teenagers with and without DLD. In response to lexico-semantic mismatches, we expected to observe N400s in our groups with and without DLD, but possibly with quantitative differences between the groups. In response to morphosyntactic mismatches, we expected our participants with TL to display more adult-like ERP patterns when processing morphosyntactic markers acquired earlier, as ERP will track behavioural abilities. In particular, we predict number agreement in NPs to elicit a more mature pattern than subject-verb agreement, as the first type is acquired earlier. Subject-verb agreement regularity might also influence ERP patterns, with regular liaison[3] eliciting more adult-like than irregular verb-final consonant marking (Marquis & Royle, 2019). We expect that participants with DLD will not achieve adult-like morphosyntactic processing, as ERPs reflecting expert patterns have not been found in any study (see Table 1).

6.2. Methods

6.2.1. Participants

A total of thirty-six pre-teens and teenagers participated in this study. The protocol was approved by the University of Montreal Research Ethics Board for educational and psychology research (CERES). All participants' parents gave written consent for their child's participation prior to the first experimental session. All participants had a hearing screen on the first day of assessment (500 Hz to 8000 Hz at 25 dB in at least one ear). Their mother tongue was French and was their language of instruction and daily use. All had normal or corrected-to-normal vision and had no history of major illnesses or prolonged hospitalization. Most of the participants were right-handed (*n* =31), as assessed using the Edinburgh Handedness Inventory French adaptation (Fromont et al., 2020).

Seventeen participants with DLD (DLD group), including 10 girls, aged between 12 and 15 years (M = 14.01; SD = 0.72) participated in this study. Most (*n* = 14) were recruited from a specialized private school for children and adolescents with learning disabilities in Montreal (Quebec, Canada) via an invitation letter sent by the school speech-language pathologist to parents of students who met selection criteria. Note that this school excludes children with disruptive behaviour, which probably explains why our group of participants with DLD includes

[3] However, note that adults show sociolinguistic variability on liaison in Québec French, see Courteau et al. (2019) for results and discussion.

more girls than boys. Other participants were recruited from a parent's association for children with DLD. All participants had a documented history of DLD, with a complete speech-language pathologist's language evaluation (including narrative and pragmatic domains) resulting in a diagnosis. All participants in the DLD group had been diagnosed before kindergarten or during the first year of primary school, and maintained significant functional impairments needing adaptations to succeed in school. These were for the most part accommodations in regular classes or enrolment in a special class, reflecting Bishop et al's (2017) definition of DLD. Note that many participants had co-morbid disorders, such as ADHD and dyspraxia. These disorders do not preclude a DLD diagnosis (see *Statement 9*; Bishop et al., 2017)). A study by Redmond Sean M. et al. (2015) shows that ADHD co-morbidity with DLD–and TL–does not increase children's errors on language assessment tasks such as sentence recall. Nevertheless, the dominant clinical profile of our DLD group was the presence of persistent language difficulties. Lastly, the group with DLD had significantly lower scores than the typical language group (TL) group (see Table 2) on the Word Classes Receptive task, which evaluates the ability to understand lexico-semantic class relationships, and Recalling sentences (both in CELF-IV$^{cnd\text{-}F}$, French version, Secord et al., 2009). The Recalling sentence task, which marshals semantic, morphological, and syntactic domains, has been shown to discriminate between typical and disordered language development in English (Conti-Ramsden et al., 2001) and in French (Courteau et al., in revision; Leclercq et al., 2014).

Nineteen participants with no history of language impairment (7 girls), aged between 8 and 14 years (M = 12.48; SD = 1.92) were included in the typical language group (TL group). Their typical developmental status was established via a questionnaire filled out during an interview with their parents, and confirmed by our linguistic and cognitive tasks. Both groups were matched on non-verbal abilities using tasks within the Cognitive Experiments IV v2 package of the Presentation® software (Version 18.0, Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com). Non-verbal working memory was assessed with the forward and backward Corsi Block tasks (Corsi, 1972) and with a delayed match-to-sample task on non-verbal stimuli (Daniel et al., 2016) with delays of 1 or 5 seconds. Participant characteristics for both groups are presented in Table 2. To compare groups statistically, we used Brunner-Munzel tests (Brunner & Munzel, 2000) as recommended by Rietveld and van Hout (2015) for group mean comparisons on skewed data with small sample sizes. Differences between groups were found in age (DLD > TL)[4] and on the Recalling sentences and Word Classes tests (DLD < TL).

[4] Participants were meant to be matched on the age variable, but recruitment was halted due to the COVID pandemic.

Table 2

Characteristics of participants

	DLD group (N = 17)		TL group (N = 19)		Brunner-Munzel tests		
	Mean	SD	Mean	SD	t_{bm}	*p*-value	CLES
Age	14.01	0.72	12.48	1.92	3.11	0.004	0.75
School	7.53	0.51	6.16	1.92	2.05	0.052	0.69
Recal	56.76	7.59	68.74	8.89	6.83	< .0001	0.12
Word Rec	12.44	4.0	16.42	3.79	3.20	< .01	0.23
Corsi–F	5.56	1.55	5.55	1.76	0.12	0.91	0.51
Corsi–B	4.94	1.06	5.60	1.76	1.18	0.25	0.39
DMTS–1s	0.88	0.10	0.89	0.11	0.49	0.63	0.55
DMTS–5s	0.84	0.13	0.82	0.15	0.24	0.81	0.48

Note. Comparisons between groups are expressed as the Brunner-Munzel statistic (t_{bm}), a p-value and a Common Language Effect Size (CLES), indicating the probability of a random observation from the DLD group being larger than a random observation from the TL group, with 0.5 being at chance. Chronological age (Age) and schooling (School) are expressed in years. Recalling sentences (Recal) and Word Classes Receptive (Word Rec) CELF-IV$^{cnd\text{-}F}$ scores are untransformed, Corsi block scores reflect forward (Corsi–F) and backward (Corsi–B) spatial spans, and delayed match-to-sample represent the accuracy for 1 second (DMTS–1s) and 5 second (DMTS–5s) delays.

6.2.2. Experimental task

Experimental tasks and stimuli were inspired by the fLEX evaluation tool, Task 4 (fLEX: Multilingual assessment of inFlectional and LEXical processing, Pourquié, 2015), where participants have to look at pictures while producing or listening to sentences involving singular and plural verb cues. A third of the pictures and verbs used in our experiment were taken from fLEX, while two-thirds were developed in order to have sufficient number for an ERP experiment. Stimuli in this study are the same as those used in (Courteau et al., 2019) with French-speaking adults. We refer to it for a detailed presentation of materials and lists development. Each participant heard 300 spoken grammatical sentences paired with a picture that matched or mismatched its morphosyntactic (n = 240) or semantic features (n = 60): half of the bimodal pairs were mismatches. Sentences began either with 1) a neutral context featuring a description of a general characteristic of the scene depicted in the picture (e.g., ‘each week’), or with 2) a subject context featuring a full NP that described the picture’s subject with lexical as well as morphosyntactic number agreement information (plural/singular, e.g., ‘the grandmother/s’). Following these contexts, verbs were presented within sentences containing third person pronouns[5], and a sentence continuation with a direct object NP, or prepositional phrase (PP, e.g., ‘in the public pool’) to avoid sentence-final effects in ERPs time-locked to verbs (Hagoort, 2003; see also Stowe et al., 2018). See Tables 3, 4 and 5 for examples of sentences.

[5] In sentences with subject contexts, the subject with an overt NP (e.g. “the girl”) was followed by a pronoun (e.g. “she”), which is grammatical (some say the pronoun is obligatory, see e.g. Auger [68]) in oral Quebec French, contrary to standard written French.

Table 3

Experimental lexico-semantic conditions and their corresponding visual stimulus.

Visual Stimulus

Sample visual stimulus presented concurrently with auditory stimuli for lexico-semantic match (1a-b) and mismatch conditions (2a-b). Note that, in addition to the mismatch at the target verb ("swims" vs. "sings"), conditions 2a-b also include a second mismatch in the prepositional phrase (here: "public pool" vs. "concert hall")

Condition	Context	Sample auditory stimuli	
Semantic Match	Neutral	(1a)	*Chaque semaine ǀ elle <u>nage</u> dans la piscine publique* 'Each week ǀ she <u>swims</u> in the public pool'
	Subject	(1b)	*La grand-mère ǀ elle nage dans la piscine publique* 'The grandmother ǀ she <u>swims</u> in the public pool'
Semantic Mismatch	Neutral	(2a)	*Chaque semaine ǀ elle !<u>chante</u> dans la salle de concert* 'Each week ǀ she !sings at a concert hall'
	Subject	(2b)	*La grand-mère ǀ elle !<u>chante</u> dans la salle de concert* 'The grandmother ǀ she !sings at a concert hall'

Note. Critical words are <u>underlined</u>. Subject= overt subject NP; ! = lexico-semantic **mismatch;** ǀ = cross-splicing point.

Table 4

Experimental morphosyntactic conditions involving consonant-final (CONS) verbs, and their corresponding visual stimuli.

Visual Stimulus

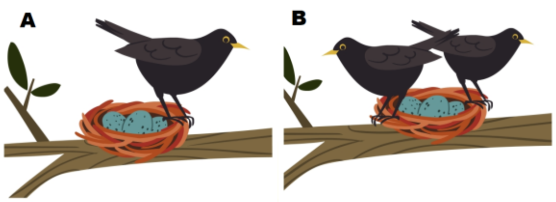


Image A: sample visual stimulus for match (1a-b) and mismatch conditions (2c-d). Image B: sample visual stimulus for match (2a-b) and mismatch conditions (1c-d).

Condition	Number	Context		Sample auditory stimuli
Morphosyntax Match	Singular	Neutral	(1a)	*Chaque printemps* ǀ *il pond dans le nid* 'Each spring ǀ he lays in the nest'
		Subject	(1b)	*Le merle* ǀ *il pond dans le nid* 'The blackbird ǀ he lays in the nest'
	Plural	Neutral	(2a)	*Chaque printemps* ǀ *ils pondent dans le nid* 'Each spring ǀ they lay in the nest'
		Subject	(2a)	*Les merles* ǀ *ils pondent dans le nid* 'The blackbirds ǀ they lay in the nest'
Morphosyntax Mismatch	Singular	Neutral	(1c)	*Chaque printemps* ǀ *il *pond dans le nid* 'Each spring ǀ he *lays in the nest'
		Subject	(1d)	**Le merle* ǀ *il *pond dans le nid* '*The blackbird ǀ he *lays in the nest'
	Plural	Neutral	(2c)	*Chaque printemps* ǀ *ils *pondent dans le nid* 'Each spring ǀ they *lay in the nest'
		Subject	(2d)	**Les merles* ǀ *ils *pondent dans le nid* '*The blackbirds ǀ they *lay in the nest'

Note. Critical words carrying agreement morphosyntactic number cues are underlined. Subject= overt subject NP; * = number mismatch; ǀ = cross-splicing point.

Table 5

Experimental morphosyntactic conditions involving liaison (LIAIS) verbs and their corresponding visual stimuli.

Visual Stimulus

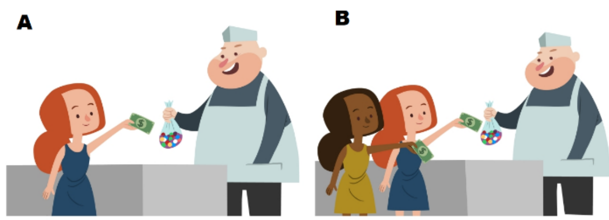


Image A: sample visual stimulus for match (1a-b) and mismatch conditions (2c-d) in the singular. Image B: sample visual stimulus for match (2a-b) and mismatch conditions (1c-d) in the plural.

Condition	Number	Context		Sample auditory stimuli
Morphosyntax Match	Singular	Neutral	(1a)	*À midi* ǀ *elle achète des bonbons au marchand* 'At noon ǀ she buys candies from the vendor'
		Subject	(1b)	*La petite fille* ǀ *elle achète des bonbons au marchand* 'The little girl ǀ she buys candies from the vendor'
	Plural	Neutral	(2a)	*À midi* ǀ *elles‿achètent des bonbons au marchand* 'At noon ǀ they buy candies from the vendor'
		Subject	(2b)	*Les petites filles* ǀ *elles‿achètent des bonbons au marchand* 'The little girls ǀ they buy candies from the vendor'
Morphosyntax Mismatch	Singular	Neutral	(1c)	*À midi* ǀ *elle *achète des bonbons au marchand* 'At noon ǀ she *buys candies from the vendor'
		Subject	(1d)	**La petite fille* ǀ *elle *achète des bonbons au marchand* '*The little girl ǀ she *buys candies from the vendor'
	Plural	Neutral	(2c)	*À midi* ǀ *elles‿ *achètent des bonbons au marchand* 'At noon ǀ they *buy candies from the vendor'
		Subject	(2d)	**Les petites filles* ǀ *elles‿ *achètent des bonbons au marchand* '*The little girls ǀ they *buy candies from the vendor'

Note. Critical words carrying agreement morphosyntactic number cues are underlined. Subject= overt subject NP; * = number mismatch; ǀ = cross-splicing point; ‿ = *liaison*

Our experiment included three types of verbs, each one related to different mismatches. Morphosyntactic conditions included two verb types with different morpho-phonological and morphosyntactic properties in the plural. First, irregular verbs where the plural is signaled by the presence of a consonant at the end of the verb[6] (consonant-final verbs, CONS, see Table 4).

[6] We used subregular and irregular verbs to assess irregular verb agreement. However, there are conflicting results in the literature about whether subregular verbs have similar productive inflection to

Second, verbs with regular subject-verb agreement signaled by the phoneme [z] resulting from liaison between the pronoun's plural form (i.e., *elles/ils* 'they.FEM/MASC' [ɛlz/ɪlz]) and the vowel onset of the following verb (liaison verbs, LIAIS, see Table 5). In the morphosyntactic incorrect conditions, the mismatches were created by using incongruent number agreement between the auditory and the visual stimuli (e.g., the sound file described two girls buying candies and the image depicted one girl buying candies). Each sentence in these conditions (CONS or LIAIS) could include either one or two number agreement mismatches: in sentence-initial neutral contexts, the mismatches occurred only on the verb, while in sentences with subject NP contexts, the mismatches occurred both within the subject context, at the determiner[7], and on the verb. See Tables 4 and 5 for examples. A third verb type was used in the lexico-semantic conditions: these had a constant phonological form in singular and plural contexts (e.g., *elle/s nage/nt* [ɛlnaʒ] 'she/they swim/s'). The lexico-semantic mismatches were created by presenting a verb (and its direct object NP or prepositional phrase) that did not match the depicted action and the general theme of the picture (e.g., the sound file described 'she swims in the public pool' and the image depicted 'she sings at the concert venue'), as illustrated in Table 3.

The experiment included 180 French verbs acquired before age 8 selected from the Manulex database (Lété et al., 2004). All verbs were matched across types on lemma frequency, age of emergence, and length (syllables and phonemes). Manulex and Lexique (New et al., 2001) were consulted to ensure that all nouns, adverbs, prepositions and adjectives used were age-

---

regular verbs or are stored as irregular verbs. See Royle et al. (2012) for a discussion. Note that in CONS verbs, vowel laxing and other irregular vowel changes could have provided hints before the end of the verb, which was the cases for 70 % of the verbs in this condition.

[7] In oral French number agreement marking is predominantly on the determiner (*le* /lœ/, *les* /lɛ/), and not on the noun, contrary to English.

appropriate and frequent. Subject grammatical gender (feminine or masculine), as well as syllable length of context phrases and of full sentences were balanced across the three verb types. For each verb, two colour drawings, with either 1 or 2 agents, were created by a professional artist. Drawings had a constant visual complexity level, avoiding superfluous or distracting details, and emphasized the action and agents described.

Sentences were recorded by a professional French-Canadian actress who read all sentences in all contexts (i.e., neutral and subject NP sentences in the singular and plural). She clearly articulated the words with natural intonation while avoiding coarticulation. Auditory stimuli were recorded at 44.1KHz in a sound attenuated booth using a Sony DAT recorder (PCM-M1, 1997). To ensure a constant voice amplitude, a sonometer was placed 10 cm in front of her mouth to monitor for deviations of +/-5Db. Using Praat software (Boersma & Weenink, 2017), we spliced the sentence-initial contexts to ensure that the neutral context for any given verb was identical in its singular and plural version, and to provide identical contexts in lexico-semantic conditions. Thus 1200 sentences were created. These were distributed throughout 4 lists resulting in 300 sentences by list. Lists were created in a counterbalanced manner where half of the sentences had neutral contexts and the other half subject NP contexts, and half of the sentences were singular and the other half plural.

6.2.3. EEG Recording

EEG was recorded continuously with a sampling rate of 500 Hz with 32 cap-mounted electrodes (WaveGuard active shielded caps, ANT, Enschede, NL) placed in accordance with the international standard 10/20 system. Electrodes used for recording cover the frontal, parietal, temporal and occipital lobes were: FP1, FP2, F3, F4, F7, F8, Fz, C3, C4, Cz, P3, P4, Pz, T7, T8, P7, P8, O1, O2, and Oz. All impedances were maintained below 5kΩ and checked prior to and after recording. EEG was amplified with an ANT Neuro eego™sports amplifier referenced to the CPz electrode.

6.2.4. Procedure

EEG recording took place in a quiet room either at the participants' school or at the *Language Acquisition and Processing Lab* (Dr. Royle director) at the University of Montreal. Upon arrival, participants provided their parent's written signed consent and underwent the audiology screen after which they were fitted with an EEG cap. Participants sat at a desk about 40 cm from a computer screen for the duration of the EEG session, for about 1 hour. Sentences were presented in an alien learning paradigm where Euzabie the alien had travelled to Quebec, was in a classroom, and had to practice French by describing pictures in a workbook. A story containing filler sentences, images, and animations was interspersed throughout the experience to maintain interest and attention. Participants listened to the spoken sentences presented binaurally via insert earphones (ER-1 Insert Earphones, Etymotic Research), while images were presented on the computer screen. A participant-controlled break was scheduled between every experimental blocks (30 items).

Participants were asked to listen to each sentence while considering all aspects of grammar and meaning, and to judge sentence acceptability as appropriately describing the simultaneously presented image, by pressing one of the two keys on a response keyboard: "acceptable" or "not acceptable". In order to avoid laterality effects, the "acceptable" button was randomly assigned to right or left sides. Both keys had a smiley face or sad face sticker so that participants did not have to memorize them. Participants were instructed to minimize movement and keep their eyes open during stimulus presentation. Six practice trials were presented at the beginning of the experiment and were excluded from further analysis. At least one researcher or assistant was present throughout the session. EEG recording was monitored during the experiment, and participants received feedback on blinking and other body movements whenever necessary, to reduce artifacts. Each trial had the following structure: a fixation cross was displayed in the centre of the screen for 1000 ms before stimulus presentation. Then the picture appeared 500 ms before the spoken sentence onset and stayed on the centre of the screen until the sentence ended. After the sentence, a blank screen appeared for 1000ms, then a response prompt ('???') appeared on the screen and remained until a button was pressed, followed by a fixation cross for 1000 ms and a blank screen for 1000 ms. After EEG recording, participants completed experimental and clinical language evaluation tasks with the first author or a research assistant. Participants attended two two-hour-and-a-half sessions on average, including two ERP recording sessions: only data from the second EEG recording session are reported here.

6.2.5. ERP measures

The offline EEG data were processed using MATLAB (MathWorks, Natick, MA, USA, v. R2018 B), EEGLAB (Delorme & Makeig, 2004, v. 2019.1), Fieldtrip (Oostenveld et al., 2011), and the ERPscope R package[8] to illustrate effects (Herbay & Steinhauer, 2020). Raw data were re-referenced to linked mastoids and filtered using Kaiser low-pass (40 Hz) and high-pass (0.3 Hz) filters. Three participants had bridged electrodes (D02, D03, D18): channel interpolations resolved these issues for two of them while one participant (D02) was excluded from further analyses. EEG signals contaminated with eye blinks were corrected using independent component analysis (ICA), as this method is an effective and objective method for correcting eye-blink artifacts in the EEG signal of young participants with language impairment

[8] http://github.com/aherbay/erpscope

(Misirliyan et al., 2020). For each participant, one component associated with eye blinks was identified with the EEGLAB plugin SASICA (Chaumon et al., 2015) and removed from the EEG signal. Movements and other artifacts were rejected using a 150µV criterion, and all uncontaminated trials were entered into the final analysis. Single-subject EEG waveforms per condition were averaged separately over 3600 ms epochs (-600 to 3000 ms), time-locked to the relevant critical word onset (underlined words in Tables 3, 4, 5), and entered into grand-average ERPs. Following artifact rejection, an average of 54/60 (SD = 5.97) trials for semantic mismatches and 210/240 trials (SD = 24) for number mismatches were analyzed per participant. All participants were retained since we rejected less then 35 % of trials per participant, the suggested criterion being 50 % for patient studies (Luck et al., 2009). We chose representative time-windows for statistical analyses based on adults' results (Courteau et al., 2019), and on previous literature on ERP components for children and adolescents (Royle & Courteau, 2014), ERP components were quantified as the mean EEG signal voltage (in µVs).

6.2.6. Analyses

In both ERP and behavioural analyses, we compared the mismatch conditions with their corresponding match ones. For instance, in the singular number mismatch sub-condition, we compared singular spoken sentences with a corresponding picture depicting one agent (match condition) to the same spoken sentence with a similar picture depicting two agents (mismatch condition). Statistical analyses for ERPs and response-times for accuracy judgments were performed using the Easy analysis and factorial experiments visualization package in R (Lawrence, MA. 2011, R package version 4.4-0 3). When degrees of freedom were above two, Greenhouse-Geisser corrections were applied to address potential violations of sphericity: in these cases, the original degrees of freedom and corrected probability levels are reported.

Regarding acceptability-judgment accuracy analyses, our score distributions were heavily skewed. Thus, we used the Aligned Rank Transform ANOVAs for R, a non-parametric approach to factorial ANOVA, which relies on a preprocessing step that aligns data before applying averaged ranks (Wobbrock et al., 2011), with *A*-scores as the dependant variable. We report the omega-squared (ω^2) as effect size, which is interpreted as follows: < 0.02, very small; 0.02 ≧≦ 0.13, small; 0.13 ≧≦ 0.26, medium; ≧ 0.26, large (Cohen, 1992). Post-hoc comparisons for interaction decomposition were done using the Wilcoxon-Mann-Withney test, and we applied the Bonferroni correction for multiples comparisons. We report the *r* statistic for effect size, which varies from 0 to close to 1. The interpretation values for *r* (Cohen, 1988) are: 0.10– 0.3 (small effect), 0.30≧≦0.5 (moderate effect) and ≧ 0.5 (large effect).

For ERP analyses, in each time-window global ANOVAs were performed separately on midline and lateral electrodes, with GROUP as a between-subject factor and the other factors as within-subject ones. For the midline channels, the semantic condition included 4 factors: CONTEXT (neutral vs. subject NP), GROUP (DLD vs. TL), CONDITION (mismatch vs. match), and ELECTRODE (Fz, Cz, Pz, and Oz). At lateral electrodes, the ANOVA included 6 factors: CONTEXT , GROUP , CONDITION , HEMISPHERE (right vs. left), ANTERIORITY (frontal, central, and posterior electrodes), and LATERALITY (lateral vs. medial). For morphosyntactic mismatch conditions, the factor NUMBER (singular vs. plural) was included for both analyses. An alpha of $p < 0.05$ was used for all statistical analyses.

Accuracy data from acceptability judgments were analyzed using detection theory for grammaticality judgment, which provides an unbiased measure of sensitivity including the participant's ability to discriminate match and mismatch conditions (Huang & Ferreira, 2020). The *A*-score (A'-score, corrected version; Zhang & Mueller, 2005) was chosen because the groups' accuracy judgments were characterised by both low and high sensitivity depending on the conditions, resulting in logistic or rectangular distributions (see appendix 3 in Macmillan & Creelman, 2005). *A*-scores of 1 reflect perfect discrimination, and 0.5 chance levels. We performed a first ANOVA with two factors, including the three VERB types (lexico-semantic, CONS, LIAIS) and GROUP (DLD vs. TL). The second one targeted morphosyntactic conditions and included four factors: CONTEXT (neutral vs. subject), NUMBER (singular, plural), VERB (CONS, LIAIS) and GROUP (DLD vs. TL). To analyze reaction times related to acceptability judgments, we filtered the data with robust estimators, a more secure method to detect outliers than means and standard deviations, as recommended by Lachaud and Renaud (2011). We used the median and the median of the absolute deviation (MAD) as estimators of central tendency and variability. We took out all reaction times that were ± 3 MAD from the median of each participant and of each 20 sub-conditions[9] to strengthen reliability. We conducted a sensitivity analysis, where global ANOVAs were performed on the unfiltered and filtered data to confirm the results. We report here the filtered data analyses. Note that all significant effects were the same in both analyses, except for one effect where visual inspection of the data revealed that filtering properly excluded outliers that were introducing biases in the dataset. We performed two ANOVAs with the same design as those used for acceptability judgments (see above).

[9] The 20 sub-conditions were as follow: Lexico-Semantics divided in 4 conditions (factors: neutral–NP context, match–mismatch); LIAIS and CONS verbs divided in 8 conditions each (factors : neutral–NP context, singular–plural, match-mismatch)

6.3. Results

6.3.1. Behavioural performance

6.3.1.1. Accuracy

Judgment accuracy *A*-scores for all conditions and both groups are listed in Table 6[10]. When comparing *A*-scores for all verb types, we found a significant main effect of GROUP ($F(1,34) = 16.94$, $p < 0.001$, $\omega^2 = 0.31$), of VERB ($F(2,68) = 40.86$, $p < 0.001$, $\omega^2 = 0.53$) and an interaction between both factors ($F(2,68) = 7.95$, $p < 0.001$, $\omega^2 = 0.16$). Interaction decomposition revealed that the TL group performed better than the DLD in sentences containing mismatches on CONS (U = 48.5, $p < 0.001$, $r = 0.58$, TL: MED = 0.92, DLD: MED = 0.85) and LIAIS verbs (U = 76, $p < 0.01$, $r = 0.45$, TL: MED = 0.94, DLD: MED = 0.85), but that both groups performed similarly when rating lexico-semantic conditions (U = 120, $p = 2.0$, r = 0.22, TL: MED = 0.96, DLD: MED = 0.95).

The ANOVA targeting morphosyntactic conditions first revealed a main effect of VERB ($F(1,34) = 8.97$, $p < 0.001$, $\omega^2 = 0.18$). This reflected the fact that in all our conditions, sentences with LIAIS verbs (MED = 0.93) generated better performance than CONS ones (MED = 0.89) in both groups. We also found main effects for the factors GROUP ($F(1,34) = 14.09$, $p < 0.001$, $\omega^2 = 0.27$), with TL (MED = 0.94) performing better than DLD (MED = 0.86), CONTEXT ($F(1,34) = 26.79$, $p < 0.001$, $\omega^2 = 0.42$), with subject context sentences (MED = 0.93) being more accurately identified than neutral ones (MED =0.87), and NUMBER ($F(1,34) = 12.54$, $p < 0.001$, $\omega^2 = 0.24$). The decomposition of a significant GROUP × NUMBER interaction ($F(1,34) = 7.67$, $p < 0.01$, $\omega^2 =$

---

[10] Refer to Section 1 of Supplementary Materials for accuracy means and standard deviations for matching and mismatching trials, and 5 supplementary figures illustrating interaction effects.

0.16) revealed no differences between plural and singular sentences in the TL group (U = 3023, $p = 0.6$, $r = 0.05$, plural: MED = 0.93, singular: MED = 0.94), while those with DLD performed more poorly on singular sentences than plural ones (U = 2834, $p < 0.05$, $r = 0.20$, singular: MED = 0.80, plural: MED = 0.88). Decomposing the GROUP × CONTEXT interaction ($F(1,34) = 10.61$, $p < 0.01$, $\omega^2 = 0.21$), we found a significant effect of context in both groups with the same pattern as found in the main effect of CONTEXT, where subject contexts were better rated than neutral ones. However, this pattern was less prominent in the TL group, with a small effect (TL: U = 2330, $p < 0.05$, $r = 0.17$, subject MED = 0.95, neutral MED = 0.92) compare to the DLD where this effect was moderate (U = 1224, $p < 0.001$, $r = 0.40$, subject MED = 0.90, neutral MED = 0.76). The significant CONTEXT × NUMBER interaction ($F(1,34) = 4.88$, $p < 0.03$, $\omega^2 = 0.10$) showed that in both plural and singular sentences, subject contexts were slightly better rated than neutral ones with significant small effects (plural: U = 1835, $p < 0.01$, $r = 0.25$, subject: MED = 0.94, neutral: MED = 0.89; singular: U = 1865, $p < 0.01$, $r = 0.26$, subject: MED = 0.92, plural: MED = 0.83).

In sum, the TL group outperformed the DLD one on LIAIS and CONS verbs, as supported by large effect sizes, while both groups performed similarly on lexico-semantic stimuli. When considering morphosyntactic conditions, *A*-scores were slightly higher for both groups on LIAIS than CONS verbs, as illustrated by a small effect size. Sentences with subject contexts were better identified than neutral ones, this effect being small in the TL group and moderate in the DLD one. Sentence number played a small role in the DLD participants' performance only in that they poorly identified singular mismatches less accurately than plural ones.

Table 6

Acceptability judgment accuracy.

Groups	DLD	Typical Language
	Mean (SD)	Mean (SD)
Lexico-semantics	0.94 (0.05)	0.96 (0.02)
Number: Consonant-final verbs	0.80 (0.10)	0.90 (0.07)
SINGULAR: **SUBJECT** CONTEXT	0.81 (0.16)	0.90 (0.11)
SINGULAR: NEUTRAL CONTEXT	0.68 (0.16)	0.85 (0.13)
PLURAL: **SUBJECT** CONTEXT	0.87 (0.14)	0.93 (0.04)
PLURAL: NEUTRAL CONTEXT	0.78 (0.15)	0.90 (0.10)
Number: Liaison verbs	0.83 (0.10)	0.91 (0.08)
SINGULAR: **SUBJECT** CONTEXT	0.85 (0.15)	0.94 (0.06)
SINGULAR: NEUTRAL CONTEXT	0.73 (0.20)	0.88 (0.15)
PLURAL: **SUBJECT** CONTEXT	0.89 (0.14)	0.94 (0.06)
PLURAL: NEUTRAL CONTEXT	0.79 (0.15)	0.89 (0.13)

Note. A-score averages (and standard deviations) for visual-auditory matching and mismatching trials in lexico-semantics, and number conditions for both consonant-final and liaison verbs morphosyntactic conditions.

6.3.1.2. Reaction times

Timing analyses were run on acceptability judgements to reveal differences in reactions between groups. No main effect or interaction involving the factor GROUP were significant. To check whether results on non-verbal tasks revealed differences between groups on reaction times, we compared groups on the Corsi block and DMTS tasks using the Brunner-Munzel test. No significant effects were found except for the Forward Corsi block task, with the DLD group being faster than the TL one ($t_{bm} = 2.18$, $p < 0.05$, CLES = 0.60). We think this difference can be explained by the fact that the DLD span on this task was lower than the TL group (see Table 2). Indeed, recalling short sequences is more likely to be faster than with long ones (Brunetti et al., 2014).

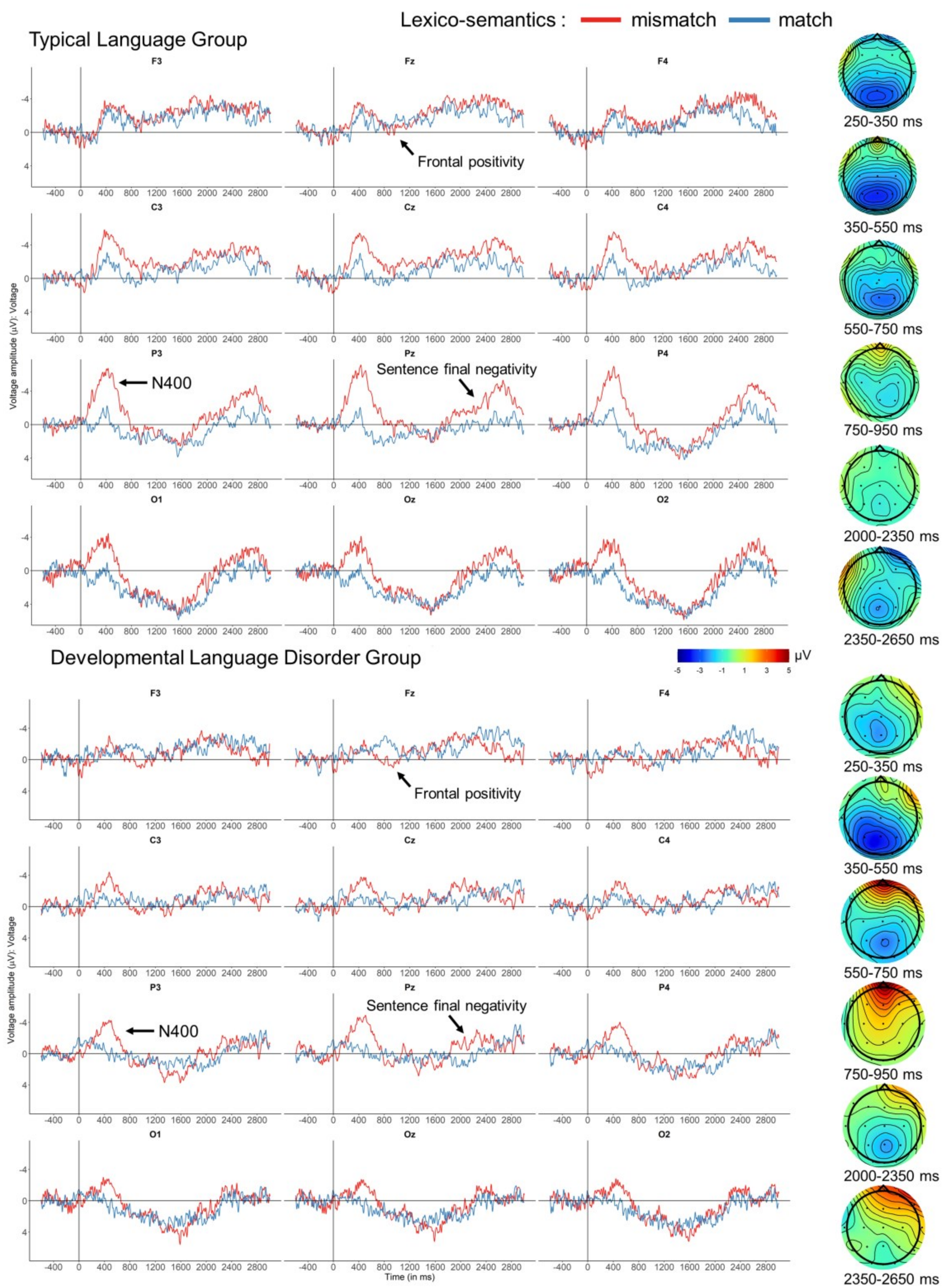
6.3.2. ERP patterns

6.3.2.1. Lexico-semantic mismatches

At verb onset, lexico-semantic mismatches elicited widely-distributed N400-like negativities over centro-parietal electrodes in both groups, when compared to the match condition (Figure 1). This negativity lasted from 250–950 ms in the TL group, and from 350–750 ms in the DLD group. It was superimposed by a positivity, which was small and restricted to frontal and midline electrodes between 750–950 ms in the TL group, and distributed in frontal and central regions for the DLD group. No effect was found between 1000 and 2000 ms, and a sentence-final negativity was elicited by mismatches from 2000–2350 ms in centro-parietal regions for both groups, but lasted longer (up to 2650 ms) in the TL group. Recall that the verbs were always followed by a sentence-final NP or a PP, and that the nouns contained in these phrases did not correspond to the picture (see Table 3). On average, verb endings were 550 ms after onset, and participants heard the NP/PP between this point and sentence offset (on average 1900 ms). Based on our stimuli's time course and observed effects, we analyzed six time-windows: 250–350 ms, 350–550 ms and 550–750 ms for early and late N400 windows, 750–950 ms for the frontal positivity, and 2000–2350 and 2350–2650 ms for the sentence-final negativity.

Figure 1

ERP effects for lexico-semantic conditions, collapsed across subject and neutral contexts.



Note. Grand-average ERPs for TL group (above) and DLD group (below) are displayed at midline and eight lateral electrodes, as well as voltage maps illustrating difference waves, time-locked to critical verb onset using a baseline of - 600 to 0 ms. Onset of the verb is indicated by the vertical calibration bar. On average the verb ended 550 ms after onset; between 600 and 1900 ms participants heard a NP/PP, which included a second lexico-semantic mismatch and ended the sentence. Compared with the correct match condition (blue line), the semantic mismatches (red) elicited a large N400 that lasted from 250 to 950 ms for the TL group, and from 350 to 750 ms in the DLD group. These were followed in both groups by a frontal positivity in frontal channels from 750-950 ms. A sentence-final negativity appeared from 2000 to 2350 ms in both groups but lasted until 2650 ms primarily in the TL group. Negative polarity is plotted upwards. Voltage maps represent difference waves (violation minus control), with negativities in blue and positivities in red.

Statistical analyses are summarized in Table 7. Significant interactions between CONDITION and topographic effects were decomposed to identify distributional patterns. The global ANOVA in the 250–350 ms time-window first revealed GROUP × CONDITION interactions in both lateral and midline channels. Decomposing these interactions confirmed an earlier N400 onset for the TL group. At lateral electrodes, interactions between GROUP, CONDITION and topographic factors revealed that the TL group's N400 was widely distributed over the right hemisphere, whereas in the left one it was significant in the more lateral electrodes only. At midline electrodes, decomposition of the ELECTRODE × CONDITION interaction showed significant shared effects at Pz and Oz in TL and DLD groups, which is related to the fact that both groups had larger negativities at posterior than frontal channels, as is illustrated in the 250–350 ms voltage maps in Figure 1. Global ANOVAs for the 350–550 ms and 550–750 ms time-windows yielded significant CONDITION effects at lateral and midline electrodes, and no interaction with GROUP, confirming the presence of a broadly distributed N400 in both groups. The CONDITION × LATERALITY, CONDITION × ELECTRODE, and CONDITION × ANTERIORITY interactions in both subsequent time-windows revealed that this effect was most prominent in posterior and occipital electrodes (350–550 ms) and in medial compared to lateral electrodes (550–740 ms).

Table 7

Global ANOVAs for lexico-semantic conditions at verb onset at time-windows of interest.

			N400		N400 and frontal positivity	Sentence-final negativity	
	df	250–350	350–550	550–750	750–950	2000–2350	2350–2650
Lateral electrodes							
Cond	(1, 33)	—	20.53***	4.26*	—	—	—
Group × Cond	(1, 33)	6.08*	—	—	3.96*	—	3.26†
DLD: Cond	(1, 15)	—	—	—	4.04†	—	—
TL: Cond	(1, 18)	25.73***	—	—	3.63†	—	—
Anteriority × Cond	(2, 66)	3.33†	12.19***	6.43**	—	4.66*	—
Anterior: Cond	(1, 34)	—	—	—	—	—	—
Central: Cond	(1, 34)	—	21.56***	6.60**	—	—	—
Posterior: Cond	(1, 34)	—	28.58***	10.08**	—	8.65**	—
Laterality × Cond	(1, 33)	7.58**	27.23***	13.86***	—	7.96**	11.03**
Lateral: Cond	(1, 34)	—	9.43**	—	—	—	—
Medial: Cond	(1, 34)	4.29*	31.28***	8.54**	—	4.52*	3.83†
Group × Hemi × Cond	(1, 33)	5.97*	—	—	—	—	6.58*
TL: Hemi × Cond	(1, 18)	6.05*	—	—	—	—	3.85†
TL: Right Hemi: Cond	(1,18)	23.27***	—	—	—	—	
Group × Lat × Cond	(1, 33)	15.31***	—	—	11.03**	—	5.97*
TL: Lat x Cond	(1, 18)	28.39***	—	—	5.18*	—	34.31***
TL: Medial: Cond	(1,18)	4.05†	—	—	5.69*	—	22.73***
TL: Lateral: Cond	(1,18)	35.64***	—	—	—	—	—
Group × Hemi × Lat × Cond	(1, 33)	6.90**	—	—	—	—	9.36**
TL: Hemi × Lat × Cond	(1,18)	6.50*	—	—	—	—	8.35**
TL: Left Hemi: Lat × Cond	(1,18)	24.03***	—	—	—	—	39.92***
TL: Left Hemi: Medial: Cond	(1,18)	18.27***	—	—	—	—	9.69**
TL: Left Hemi: Lateral: Cond	(1,18)	—	—	—	—	—	3.53†
Group × Ant × Hemi × Cond	(2, 66)	3.86†	—	—	—	—	9.47**
TL: Ant × Hemi × Cond	(1,18)	—	—	—	—	—	9.53***
TL: Frontal: Hemi × Cond	(1,18)	—	—	—	—	—	6.53*
TL: Frontal: Right Hemi: Cond	(1,18)	—	—	—	—	—	8.46**
TL: Central: Hemi × Cond	(1,18)	—	—	—	—	—	4.81*
TL: Central: Right Hemi: Cond	(1,18)	—	—	—	—	—	6.83*
Midline electrodes							
Cond	(1, 33)	4.91*	33.94***	8.88**	—	6.49*	7.33**
Group × Cond	(1, 33)	10.59*	—	—	7.87**	—	6.84**
DLD: Cond	(1, 15)	—	—	—	4.04†	—	—
TL: Cond	(1, 18)	31.45***	—	—	3.63†	—	35.29***
Electrode × Cond	(3, 99)	5.93**	11.13***	12.44***	5.15*	4.21*	4.92*
Fz: Cond	(1, 34)	—	—	—	4.47*	—	—
Cz: Cond	(1, 34)	—	24.57***	9.17**	—	3.81†	3.79†
Pz: Cond	(1, 34)	9.86**	48.41***	21.33***	—	14.13***	14.61***
Oz: Cond	(1, 34)	9.03**	18.69***	9.66**	—	9.55**	8.87**

Note. Only significant results and trends are presented. Cond = Condition; Ant = Anteriority; Lat = Laterality; Hemi = Hemisphere. †: $p < 0.10$, *: $p < 0.05$, **: $p < 0.01$, and ***: $p < 0.001$.

In the 750–950 ms time-window, the Group × Condition × Laterality interaction and its decomposition revealed a significant effect of Condition for the TL group, confirming that the negativity was still present for these participants with a focus in medial electrodes. At midline electrodes, the decomposition of Group × Condition did not reveal any significant effect. Decomposition of the ELECTRODE × CONDITION interaction in the midline revealed a significant effect of CONDITION at Fz, suggesting that a small frontal positivity was elicited in both groups. Global ANOVAs for the sentence-final negativity supported the interpretation that, in the 2000–2350 ms time-window, both groups elicited an effect. Significant interactions for CONDITION × ANTERIORITY and CONDITION × LATERALITY showed that the sentence-final negativity emerged in medial and posterior regions for the lateral electrodes, as well as midline electrodes as shown by the ELECTRODE × CONDITION interaction, with effects driven by PZ and OZ. This negativity continued from 2350 to 2650 ms primarily in the TL group. In lateral electrodes, the interactions including GROUP, CONDITION and topographical factors indicated that the TL negativity was maximal in medial left hemisphere channels. In the midline, the GROUP × CONDITION interaction decomposition yielded a significant effect in the TL group only. The ELECTRODE × CONDITION interaction revealed maximal effects at PZ and Oz, which reflects the fact that posterior channels remained negative in both groups, as seen in the 2350–2650 ms voltage maps in Figure 1. For all the time-windows, the factor CONTEXT did not result in any significant main effects or interactions involving CONDITION or GROUP.

In short, lexico-semantic mismatches elicited N400s in both groups, but with an earlier onset and a later offset for the TL group. This was followed by a small frontal positivity restricted to Fz and a sentence-final negativity in both groups. This last effect was at first broadly distributed over the scalp in both groups. From 2350–2650 ms, it continued to be strong in the TL group but was restricted to Pz and Oz in the DLD group. Note that this late time-window is more prone to be contaminated by components related to button pressing (e.g., see the CONS singular verb section), and these effects will accordingly be interpreted with caution.

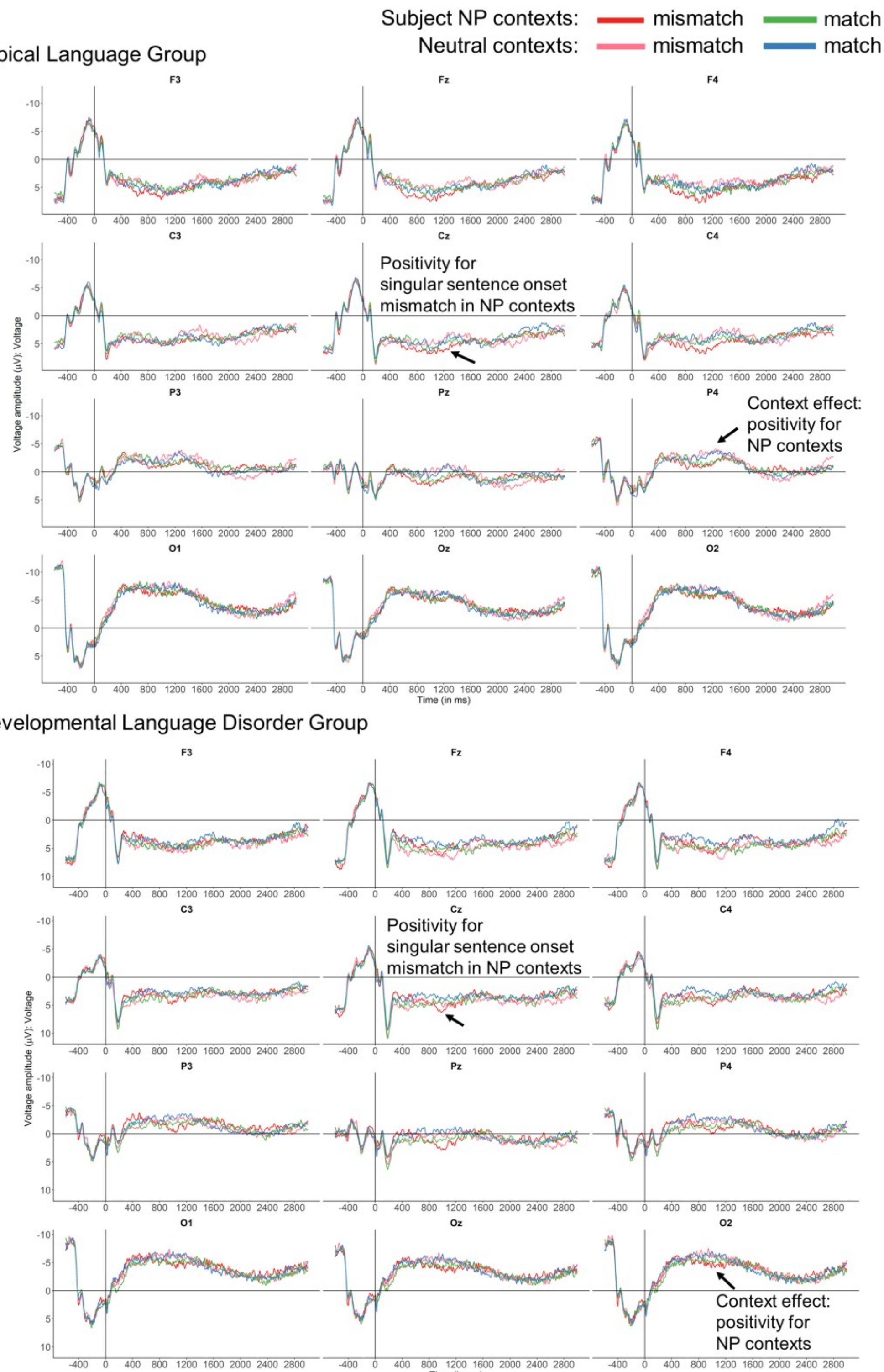
6.3.2.2. ERPs effects for number mismatches at sentence onset

We will now turn to number-agreement mismatches, which were created by combining singular or plural visually presented agents with mismatching auditory number cues. This was done using either the determiner at sentence onset, described in the present section, or the verb further downstream in the sentence, which we will present in the following sections.

Recall that participants heard two kinds of contexts at sentence onset: neutral contexts (e.g., ‘each spring’) or subject contexts (e.g., ‘the blackbird/s’). The distinction between our two kinds of verbs, CONS and LIAIS, did not play a role at this point of the sentence, so we present ERPs for both conditions combined. Sentence onset ERPs for both groups (see Figure 2) show that the first 900 ms (-600–300 ms) are dominated by sensory components. Visual-onset components related to picture presentation at -500 ms can be seen in posterior and occipital electrodes, and auditory-onset components elicited by the spoken sentence onset, starting at 0 ms, are elicited in fronto-central ones. We compared mismatch and match conditions for both subject and neutral contexts, collapsed across singular and plural sub-conditions (Figure 2). In the DLD group, ERPs revealed a negativity in the 400–600 ms time-window for the mismatch condition in subject contexts. This same condition also elicited a broadly distributed positivity in both groups, starting around 600 ms for the TL group and 800 ms for the DLD group. We will see in the statistical analyses that this P600-like positivity was driven by the singular sub-condition only, as illustrated in Figure 3. Starting around 1000 ms, we can clearly see context effects in posterior and occipital channels in both groups, where the match and mismatch conditions in subject contexts are more positive than neutral ones. Following visual inspection, we ran Global ANOVAs on 400–600 ms, 600–800 ms, 800–1000 ms and 1000-1200 ms time-windows.

Figure 2

ERP effects of context mismatches at sentence onset.



Note. Grand-average ERPs for the TL group (above) and the DLD group (below) are displayed at midline and eight lateral electrodes, time-locked to sentence onset (vertical bar) using a baseline of - 600 to 0 ms. When compared with the subject context correct condition (green), correct neutral contexts (blue) and mismatch (pink) conditions, the subject context mismatch condition (red) elicited a broadly distributed P600-like positivity in both groups from 600 to 1000 ms, which was driven by the singular subject context sub-conditions. From 600 to 1200 ms, both correct (green) and mismatch (red) subject contexts elicited a positivity in both groups, when compared to neutral contexts (blue, pink). This positivity migrated to more posterior and occipital regions after 1000 ms. The two neutral context conditions will be disambiguated further downstream at the verb and are not yet different at sentence onset.

Statistical analyses are summarized in Table 8. Between 400–600 ms, no main effect of CONDITION was found, and significant interactions were present in the DLD group only. In lateral channels, the GROUP × CONTEXT × CONDITION interaction decomposition revealed an unexpected significant effect of CONDITION for subject contexts. The same decomposition for the midline channels revealed a significant effect of CONDITION for neutral contexts only, which did not, however, feature any linguistic (or visual) difference between match and mismatch conditions at this point. The fact that the neutral context effect was not driven by our stimuli and that none of these effects were found in the TL group nor the adults (Courteau et al, 2019) prompted us to check if this effect was shared by a majority of the DLD participants. We found that both interactions were mainly driven by three participants, and these two significant interactions disappeared when they were taken out of the grand average (see Supplementary Materials Section 2 for details). Therefore, we did not consider this time-window's results to be reliable and we refrain from interpreting it.

Table 8

Global ANOVAs for sentence onset conditions at time-windows of interest.

		Early effects		Positivity	
	df	400–600	600–800	800–1000	1000–1200
LATERAL ELECTRODES					
CONDITION	(1,33)	—	—	—	—
CONTEXT	(1,33)	—	4.13*	6.57*	—
ANTERIORITY × CONT	(2,66)	—	—	2.33†	12.23***
POSTERIOR: CONT	(1,44)	—	—	—	17.68***
LATERALITY × CONT	(1,33)	—	—	3.79†	4.26*
MEDIAL: CONT		—	—	—	8.65**
LAT × ANT × CONT	(2,66)	—	—	—	4.28*
CENTRAL: LAT × CONTEXT	(1,34)	—	—	—	9.21**
CENTRAL: MEDIAL: CONTEXT	(1,34)	—	—	—	2.90†
POSTERIOR: LAT × CONTEXT	(1,34)	—	—	—	8.95**
POSTERIOR: LATERAL: CONTEXT	(1,34)	—	—	—	14.54***
POSTERIOR: MEDIAL: CONTEXT	(1,34)	—	—	—	17.60***
HEMI × CONT	(1,33)	3.45†	—	4.60*	3.82†
RIGHT HEMI: CONT	(1,34)	—	—	9.77**	—
LEFT HEMI: CONT	(1,34)	—	—	3.30†	—
GROUP × HEMI × CONT	(1,33)	—	—	—	6.74*
TL: HEMI × CONT	(1,18)	—	—	—	14.65**
TL: RIGHT HEMI: CONT	(1,18)	—	—	—	7.52**
GROUP × CONT × COND	(1,33)	7.28**	—	—	—
DLD: CONT × COND	(1,15)	7.13*	—	—	—
DLD: NP: COND	(1,15)	4.44*	—	—	—
HEMI × GROUP × CONT × COND	(1,33)	—	—	—	4.78*
RIGHT HEMI: GROUP × CONT × COND	(1,33)	—	—	—	5.42*
RIGHT HEMI: NEUT: GROUP × COND	(1,33)	—	—	—	6.58*
RIGHT HEMI: NEUT: DLD: COND		—	—	—	6.84*
HEMI × GROUP × ANT × CONT × COND	(2,66)	—	—	—	4.40*
RIGHT HEMI: GROUP × ANT × CONT × COND	(2,66)	—	—	—	4.53*
RIGHT HEMI: FRONTAL: GROUP × CONT × COND	(1,33)	—	—	—	6.73*
RIGHT HEMI: FRONTAL: TL: CONT × COND	(1,18)	—	—	—	4.84*
RIGHT HEMI: FRONTAL: TL: NP: COND	(1,18)	—	—	—	4.58*
RIGHT HEMI: CENTRAL: GROUP × CONT × COND	(1,33)	—	—	—	5.09*
RIGHT HEMI: CENTRAL: TL: CONT × COND	(1,18)	—	—	—	3.85†
GROUP × LAT × CONT × COND	(1,33)	—	6.65*	—	5.42*
MEDIAL: GROUP × CONT × COND	(1,33)	—	—	—	2.92†
CONT × NUMBER × COND	(1,33)	—	5.00*	6.47*	—
NP: NUMBER × COND	(1,34)	—	9.74**	12.39***	—
NP: SG: COND	(1,34)	—	7.03**	14.83***	—
ANT × CONT × NUMBER × COND	(2,66)	—	—	—	5.15*
FRONTAL: NUMBER × COND × CONT	(1,34)	—	—	—	4.52*
FRONTAL: NP: NUMBER × COND	(1,34)	—	—	—	5.09*
FRONTAL: NP: SG: COND	(1,34)	—	—	—	7.90**
LAT × CONT × NUMBER × COND	(1,33)	—	—	9.24**	—
MEDIAL: CONT × NUMBER × COND	(1,34)	—	—	7.70**	—
MEDIAL: NP: NUMBER × COND	(1,34)	—	—	16.48***	—
MEDIAL: NP: SG : COND	(1,34)	—	—	17.75***	—
CONT × LAT × ANT × NUMBER × COND	(2,66)	5.99**	8.83**	—	—
NP: LAT × ANT × NUMBER × COND	(2,68)	—	3.22†	—	—
GROUP × CONT × NUMBER × COND	(1,33)	—	—	—	4.70*
TL: CONT × NUMBER × COND	(1,18)	—	—	—	4.60*
TL: NP: NUMBER × COND	(1,18)	—	—	—	5.22*
TL: NP: SG: COND	(1,18)	—	—	—	21.76***
MIDLINE ELECTRODES					
CONDITION	(1,33)	—	—		
CONTEXT	(1,33)	—	—	9.59**	8.60**
CONTEXT × ELECTRODE	(3,99)	—	—	—	7.79*
CZ: CONTEXT	(1,34)	—	—	—	2.97†
PZ: CONTEXT	(1,34)	—	—	—	14.66***

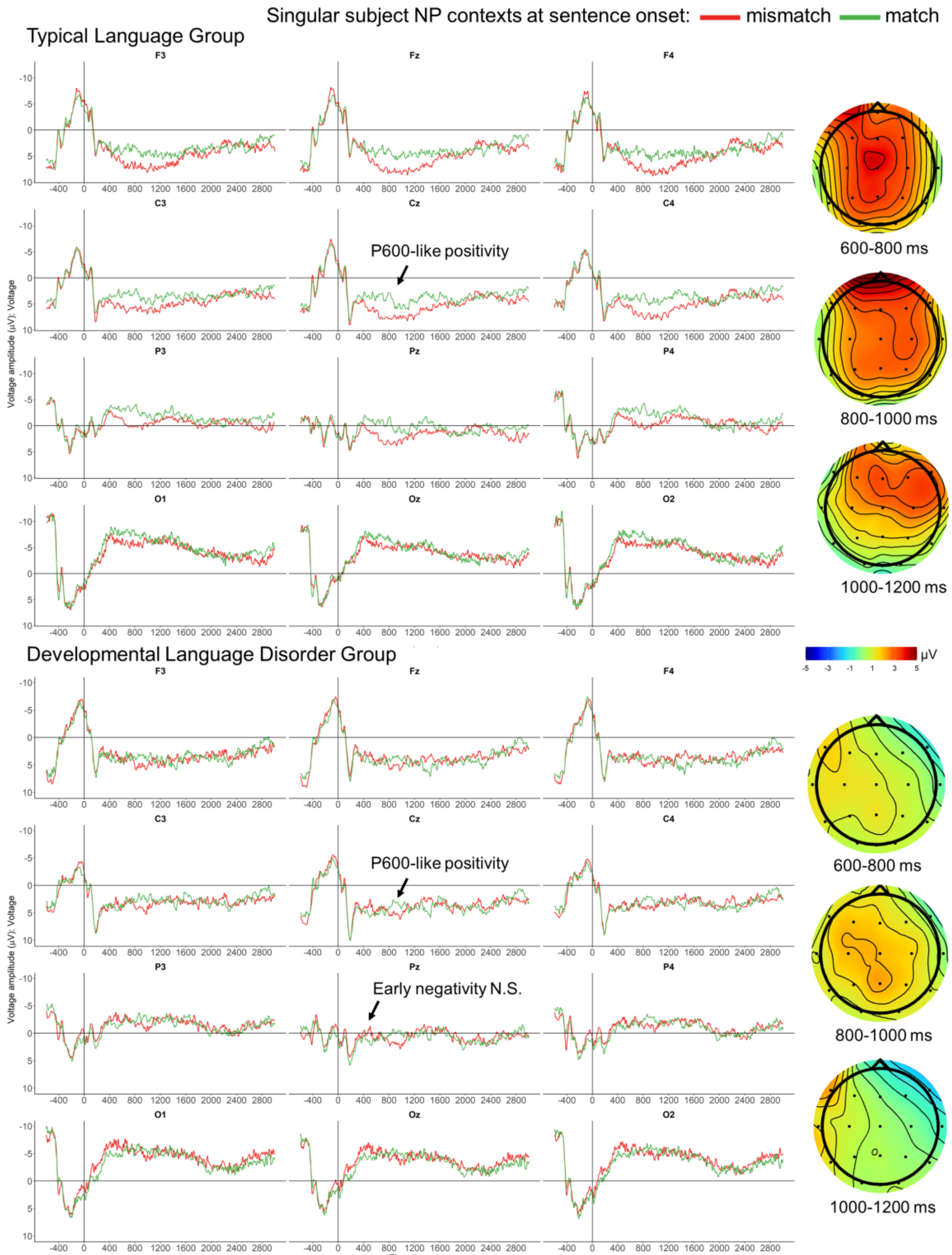
Oz: Context	(1,34)	—	—	—	15.95***
Group × Cont × Cond	(1,33)	6.82**	5.21*	—	—
DLD: Cont × Cond	(1,15)	9.07**	—	—	—
DLD: NP: Cond	(1,15)	3.96†	—	—	—
DLD: Neutral: Cond	(1,15)	6.57*	—	—	—
TL: Cont × Cond	(1,18)	—	3.75†	—	—
Electrode × Cont × Group × Cond	(3,99)	—	3.30*	—	4.57*
Fz: Cont × Group × Cond	(1,33)	—	—	—	8.99*
Fz: Neut: Group × Cond	(1,33)	—	3.52 t	—	5.38*
Fz: Neut: DLD : Cond	(1,15)	—	—	—	4.59*
Fz: NP: Group × Cond	(1,33)	—	6,13*	—	—
Fz: NP: TL : Cond	(1,18)	—	5.80*	—	—
Cont × Number × Cond	(1,33)	—	6.26*	8.08**	—
NP: Number × Cond	(1,34)	—	11.80***	12.30***	—
NP: SG: Cond	(1,34)	—	8.97**	14.14***	—
Cont × Number × Electrode × Cond	(3,99)	—	5.12**	2.70†	3.64*
NP: Number × Elect × Cond	(3,102)	—	3.04*	—	3.48*
NP: SG: Elect × Cond	(3,102)	—	—	—	2.47†

Note. Only significant results and trends are presented. Cond = Condition; Cont = Context; NP = Subject NP context; Neut = Neutral context; Numb = Number; SG = Singular; Ant = Anteriority; Lat = Laterality; Hemi = Hemisphere. †: $p < 0.10$, *: $p < 0.05$, **: $p < 0.01$, and ***: $p < 0.001$.

In the Global ANOVAs for the 600–800 ms and the 800–1000 ms time-windows, first we found significant main effects of Context showing that the two contexts were processed differently by our participants. The subject contexts elicited a positivity that was at first broadly distributed and became right-lateralized between 800 and 1000 ms as revealed by the significant Hemisphere × Context interaction. Furthermore, we found significant effects of Condition shared by both groups for singular sentences in subject contexts. These effects confirmed the broadly distributed P600-like positivity for the singular mismatch condition for both groups, as seen in Figure 3. Indeed, Context × Number × Condition interaction decompositions were significant for lateral and midline channels. This positivity was at first widely distributed over the scalp and then became focalized in more medial lateral electrodes in the 800–1000 ms time-window as indicated by the decomposition of the Context × Laterality × Number × Condition interaction. Lastly, we found an effect restricted to the Fz electrode between 600–800 ms for the TL group only, revealing a Condition effect supporting a positivity for both singular and plural sentences NP contexts, as indicated by the Group × Context × Electrode × Condition interaction decomposition.

Figure 3

ERP effects for singular subject NP context sub-conditions, at sentence onset.



Note. Grand-average ERPs for the TL group (above) and the DLD group (below) are displayed at midline and eight lateral electrodes, as well as voltage maps illustrating the difference waves, time-locked to sentence onset (vertical bar) using a baseline of -600 to 0 ms. Compared to singular subject contexts in the match condition (green), mismatches elicited a P600-like positivity in both groups. In the TL group, this positivity was present from 600 to 1200 ms, while in the DLD group, it increased from 600 to 800 ms and peaked between 800 and 1000 ms. These distinct timelines for each group are illustrated in the voltage maps representing difference waves (mismatch minus match).

In the 1000–1200 ms time-window, statistical analyses revealed a somewhat less homogenous pattern. First, subject contexts when compared to neutral ones elicited a positivity distributed in posterior and occipital channels in both participant groups, as illustrated in Figure 2. As expected, decomposition of interactions involving CONTEXT, ANTERIORITY and LATERALITY showed a significant difference between subject and neutral contexts in posterior medial and more external lateral channels. This effect of CONTEXT was more prominent in the right hemisphere for the TL group, as supported by the significant GROUP × HEMISPHERE × CONTEXT interaction decomposition. A similar posterior pattern was found in midline channels, where the CONTEXT × ELECTRODE decomposition revealed significant effects at Pz and OZ in both groups. Regarding CONDITION effects for singular sentences in subject contexts, the positivity was still present for the TL group in lateral channels only, as indicated by the decomposition of the GROUP × CONTEXT × NUMBER × CONDITION interaction. Decomposition of the 4-way interaction including ANTERIORITY, CONTEXT, NUMBER and CONDITION factors pointed towards the fact that both groups were still exhibiting a frontal positivity for singular sentences in subject contexts However, inspecting the 1000–1200 ms voltage map in both groups (Figure 3), we believe that this positivity was primarily driven by the TL group only, in which this effect lasted beyond 1200 ms. We also found a 5-way interaction involving the TL group, although this time indicating that the frontal positivity for mismatches in subject context, regardless of sentences number, was now right-lateralized in frontal channels. The DLD group showed significant effects of condition for the neutral context, as indicated by the GROUP ×

HEMISPHERE × CONTEXT × CONDITION and GROUP × CONTEXT × ELECTRODE × CONDITION interactions in lateral and midline channels respectively. We found that the two interactions were primarily driven by 3 participants, and both interactions were not significant when the 3 participants were removed from the grand average (see Supplementary Materials section 2 for details) : this result will not be interpreted further.

Difference between groups, in 600–800 ms and 800–1000 ms time-windows.

Both the analyses in the 600–800 and 800–1000 ms time-windows revealed a broadly distributed positivity for singular sentences in subject contexts across TL and DLD groups. When looking at the topographic maps in Figure 3, the time-course of this positivity seems different in each group: for TL participants, the positivity is largest from 600–800 ms and slightly decreases in the following time-window, whereas in the DLD group, the positivity emerges in the 600–800 ms time-window and seems to reach its maximum amplitude between 800 and 1000 ms. To test this assumption statistically, we ran ANOVAs comparing the two time-windows (i.e., including the additional factor TIME-WINDOW) at midline channels, on the mean amplitude of the positivity (i.e., the *difference wave* for mismatch minus match sentences). As expected, results indicated a main effect of GROUP ($F(1,33) = 4.69$, $p < 0.05$), which interacted with TIME-WINDOW ($F(1,33) = 12.45$, $p < 0.001$). First, we decomposed this interaction by GROUP, and found that in the TL group no difference was found between the two time-windows, confirming that the positivity was constant (TL: TIME-WINDOW, $F(1,18) = 2.10$, $p = 0.17$, 600–800 ms: M = 3.21µV, 800–1000 ms: M = 2.38µV). In contrast, the DLD group showed a significant TIME-WINDOW effect (DLD: $F(1,15) = 14.27$, $p < 0.001$), reflecting a substantial increase of the positivity from the first to the second time-window (600–800 ms: M = -0.23µV, 800–1000 ms: M = 1.69µV). The growing amplitude across the time-windows in the DLD group is also confirmed when decomposing by TIME-WINDOW. Indeed, we found a significant difference between groups (i.e., a larger amplitude for TL participants) in the 600–800 time-window ($F(1,33) = 13.36$, $p < 0.0001$), but not in the 800–1000 time-window ($F(1,33) = 0.38$, $p = 0.54$). To sum up, these results showed that both groups elicited a similar positivity in response to mismatches on singular CONS verbs, but with a slower onset and a smaller amplitude for the DLD group whose P600 peaked later than in the TL group.

Overall, three leading effects were found in the sentence onset condition: 1) A context effect in both groups, where subject contexts were more positive than neutral ones from 600 to 1200 ms in all electrodes. From 1000 ms, this positivity migrated to more posterior and occipital regions. 2) A broadly distributed P600-like positivity in both groups in response to singular sentence mismatches in subject contexts from 600 to 1000 ms. In the TL group, the positivity was present from 600 to 1200 ms, and in the DLD group, it increased from 600 to 800 ms and peaked between 800-1000 ms. 3) In the TL group this singular sentence-related positivity lasted until 1200 ms, but only at frontal electrode sites.

6.3.2.3. ERPs for Number Mismatches on Verbs

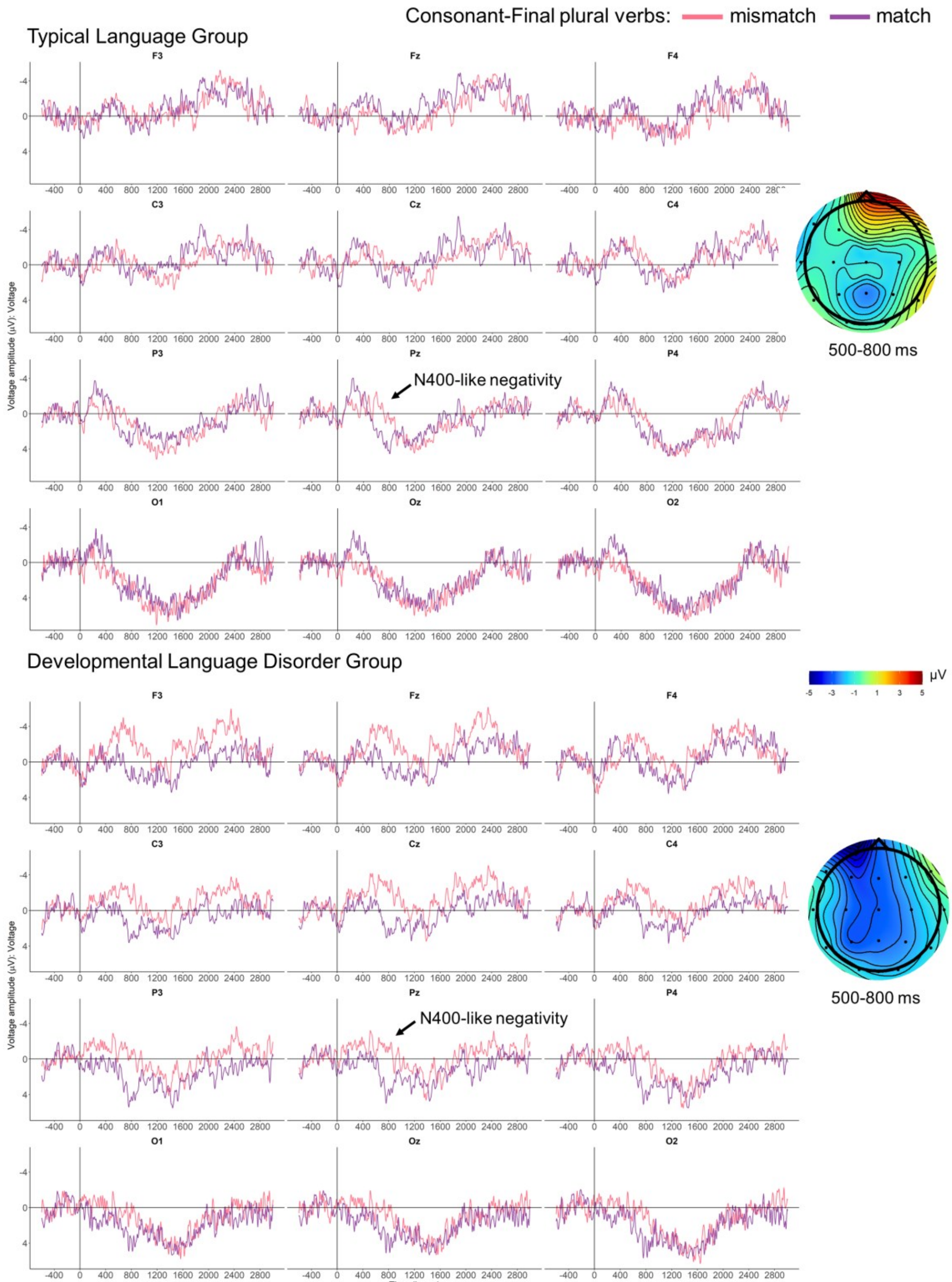
We will now report mismatch effects at target verbs in neutral context sentences since these are the ones presenting a number cue for the first time at the verb-target. At sentence onset, CONS and LIAS conditions were indistinguishable, but downstream at the target verb they differed. For CONS verbs, the verb-agreement number cue is available only at the verb-final phoneme (e.g., *ils pondent* [ilpɔ̃d] 'they lay'), and ERPs are time-locked to verb onset, while for LIAIS verbs, this information is already available at pronoun offset (i.e., *elles/ils* 'they.FEM/MASC' [ɛlz/ɪlz]), which is why ERPs in this condition are time-locked to pronoun onset to avoid ERP baseline issues. We will first focus on CONS verbs and then turn to LIAIS ones. As in Courteau et al. (2019), plural and singular sub-conditions are analysed separately, since in number mismatches on verbs they represent two different kinds of conditions: omission for singular verbs and commission for plural verbs.

6.3.2.3.1. ERPs for Consonant-Final plural conditions

As seen in Figure 4, visual inspection of the waveforms suggests that relative to the match condition, mismatches elicit early effects between 200–400 ms in both groups: a widely distributed positivity for the TL group and a posterior negativity for the DLD group. This was followed in both groups by an N400-like negativity between 500–800 ms, which is limited to centro-parietal electrodes in the TL group and broadly distributed in the DLD group. This time-window could be considered as rather late for an N400, but recall that in this condition the morphosyntactic cue revealing plural number is a verb-final consonant. On average, consonant-final onset was 275 ms following the verb's onset, which explains the delay. Between 1200–1500 ms, we observed a centro-parietal positivity resembling a P600 in the TL group and a fronto-central negativity for the DLD group. Based on these observations, three time-windows were analysed: 200–400 ms, 500–800 ms, and 1200–1500 ms.

Figure 4

ERP number mismatch effects for consonant-final plural verbs in neutral contexts.



Note. Grand-average ERPs for the TL group (above) and the DLD group (below) are displayed at midline and eight lateral electrodes, as well as with voltage maps illustrating difference waves time-locked to CONS verb onset (vertical bar) using a baseline of -600 to 0 ms. In both groups, plural mismatches (pink) elicited a broadly distributed N400-like negativity between 500 and 800 ms compared to the match condition (purple). The fronto-central positivity that can be seen in the TL group central electrodes between 1200 and 1500 ms did not reach significance.

Statistical analyses are summarized in Table 9. The global ANOVA for the early effects in the 200–400 ms time-window did not reveal any significant effects or interactions involving CONDITION at either midline or lateral electrodes. Between 500–800 ms, a significant effect of CONDITION in both lateral and midline electrodes was found, thus confirming the presence of an N400-like negativity in both groups. The effect was more noticeable in the medial lateral electrodes when compared to more lateral electrodes, as supported by a significant CONDITION × LATERALITY interaction. In the 1200–1500 ms time-window, the only significant effect was an interaction with GROUP × ANTERIORITY × CONDITION for the lateral electrodes, which did not reveal significant effects when decomposed. This means that the centro-parietal positivity resembling a P600 for the TL group did not reach significance, neither did the apparent fronto-central negativity for the DLD group.

Table 9

Global ANOVAs for consonant-final plural verbs at verb onset at time-windows of interest

	df	200–400	N400-like Negativity 500–800	1200–1500
LATERAL ELECTRODES				
CONDITION	(1,33)	—	4.17*	—
GROUP		—	—	—
LATERALITY × COND	(1,33)	—	9.47*	—
LATERAL: COND	(1,34)	—	—	—
MEDIAL: COND	(1,34)	—	8.23**	—
GROUP × LAT × COND	(1,33)	—	—	5.83*
TL: LAT × COND	(1,18)	—	—	4.72*
GROUP × ANT × LAT × COND	(2,66)	—	3.39*	—
FRONTAL: GROUP × LAT ×COND	(1,33)	—	4.11†	—
MIDLINE ELECTRODES				
CONDITION	(1,33)	—	10.08**	—
GROUP	(1,33)	—	—	—

Note. Only significant results and trends are presented. Cond = Condition; Ant = Anteriority; Lat = Laterality; Hemi = Hemisphere. †: $p < 0.10$, *: $p < 0.05$, **: $p < 0.01$, and ***: $p < 0.001$.

6.3.2.3.2. Response contingency analyses and ERP amplitude for CONS plural conditions

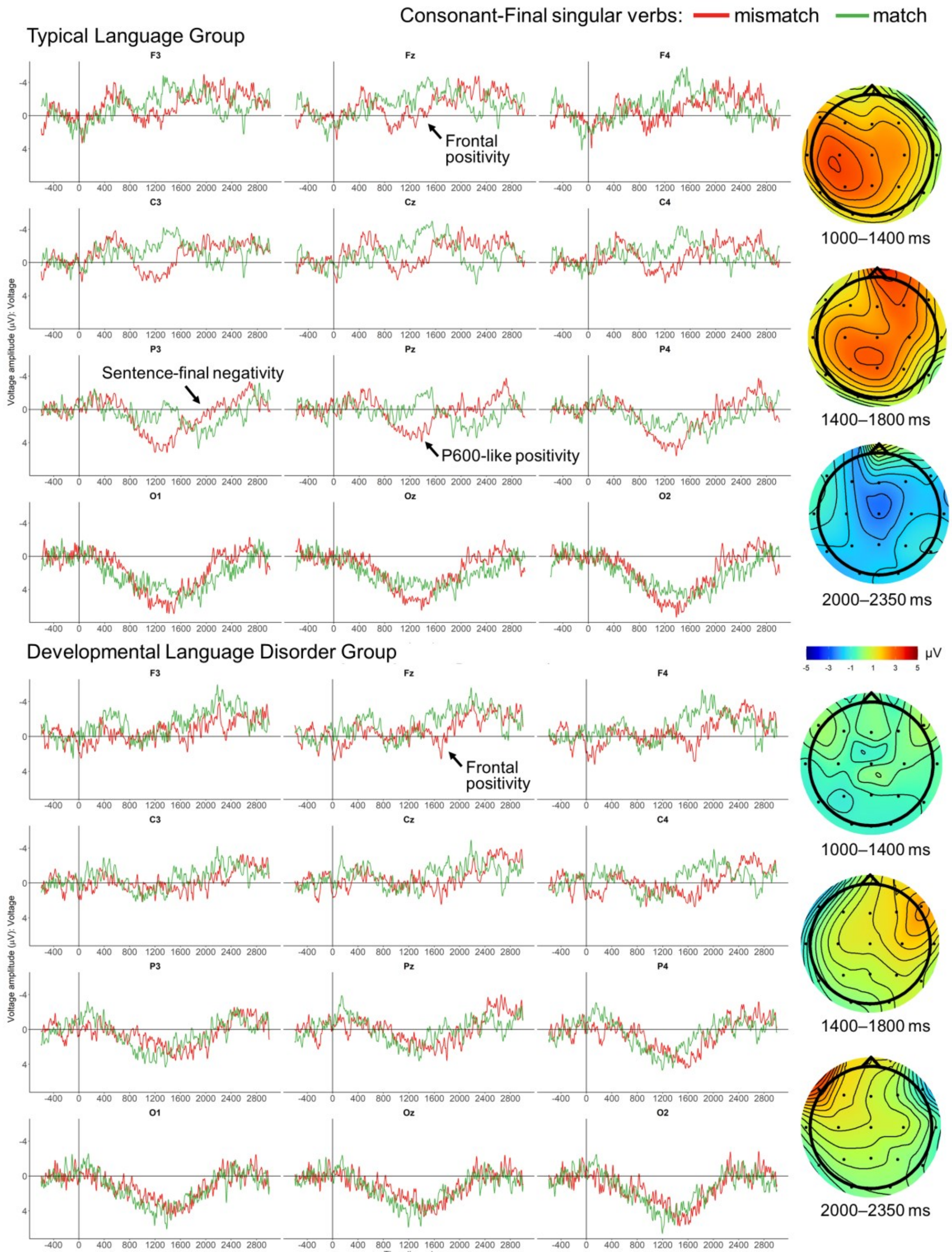
The CONS plural condition elicited an N400-like negativity in the 500–800 ms time-window in both groups. This result raised the question as to whether participants' ability to distinguish mismatches from matches was directly reflected by ERP measures. Indeed, we know that children and teenagers demonstrate considerable variability in their behaviour, and this could be reflected in their ability to process and identify mismatches. We investigated this relationship in two ways, in trial-based and participant-based analyses. In trial-based analyses, single-subject ERP averages can be "response contingent" that is based on correct behavioural trials, or on "all trials" irrespective of accuracy (White et al., 2012). Response-contingent analyses revealed an N400-like negativity, that had more of a centro-occipital distribution than the all-trials effect, but which did not differ in terms of amplitude. In participant-based analyses, we did not find significant correlations between their ability to detect mismatches and ERP amplitude, nor with their age or their sentence-repetition score. Considering that we were not able to demonstrate a relationship between participants' ability to discriminate mismatch from match CONS plural sentences and ERP measures in trial- or participant-based analyses, we did not pursue these further in the following conditions. Refer to Section 3 of Supplementary Materials for detailed analyses.

6.3.2.3.3. ERPs for CONS singular conditions

As illustrated in Figure 5, CONS singular mismatches elicited rather different ERP patterns for each group. In the TL group, mismatches elicited a negativity (500–700 ms) in central electrodes near the midline, followed by a large P600-like positivity with a posterior maximum from 1000–1400 ms, which spread to frontal channels between 1400–1800 ms. Again, this could be considered late for a time-window reflecting a P600 but recall that participants could only have processed the singular information after the verb end, when omission of the plural number cue becomes apparent. In the TL group, we also observed a widely-distributed negativity from 2000 to 2700 ms, suggesting a sentence-final negativity. For the DLD group, mismatches elicited a small early positivity in all channels (100–300 ms), followed by a fronto-central right-lateralized positivity (1400–1800 ms). From 2400–2700 ms, we found a small negativity in central and posterior channels. Note that late effects that are present after 2400 ms are expected to have been highly contaminated by motor components related to the judgment task because all conditions had ended by this time (M =1760 ms, SD = 310 ms, MIN = 1160 ms, MAX = 2390 ms). Considering this, we did not run analyses after 2400 ms. The following time-windows were thus analyzed: 100–300 ms, 500–700 ms, 1000–1400ms, 1400–1800ms, 2000–2350 ms.

Figure 5

ERP number mismatch effects for consonant-final singular verbs in neutral contexts.



Note. Grand-average ERPs for the TL group (above) and the DLD group (below) are displayed at midline and eight lateral electrodes, as well as voltage maps illustrating difference waves, time-locked to CONS verb onset (vertical bar) using a baseline of -600 to 0 ms. Compared to the match condition (green), singular mismatches (red) elicited a P600-like positivity in the TL group from 1000 to 1400 ms, and a right-lateralized frontal positivity in both groups from 1400 to 1800 ms. In the TL group only, we observe a sentence-final negativity between 2000 and 2350 ms.

We ran global ANOVAs (see Table 10) for early effects in both groups, and neither the 100–300 ms nor the 500–700 ms time-windows revealed significant main effects of CONDITION nor interactions with GROUP or topographical factors. Two time-windows were selected to assess the positivity in both groups, 1000–1400 ms and 1400–800 ms. The ANOVAs on the 1000–1400 ms time-window revealed significant interactions with GROUP and CONDITION in both lateral and midline electrodes. Decomposition of these interactions confirmed a broadly distributed P600-like positivity for the TL group only. In the 1400–1800 ms window we found main effects of CONDITION in both lateral and midline electrodes. In the lateral electrodes, interactions between CONDITION and topographic factors ANTERIORITY, HEMISPHERE and LATERALITY revealed that this positivity was prominent in frontal right-hemisphere electrodes, and in more medial lateral electrodes. Lastly, the 2000–2350 ms time-window revealed significant GROUP × CONDITION interactions in lateral and midline channels. Decomposition of these interactions confirmed a sentence-final negativity for the TL group in midline channels, and a trend in lateral channels.

Table 10

Global ANOVAs for consonant-final singular verbs at verb onset at time-windows of interest

		Early effects		P600-like positivity	Frontal positivity	Sentence-final negativity
	df	100–300	500–700	1000–1400	1400–1800	2000–2350
Lateral electrodes						
Condition	(1,33)	—	—	—	3.45*	—
Group		—	—	—	—	—
Group × Cond	(1,33)	—	—	5.58*		4.80*
DLD: Cond	(1,15)	—	—	—		—
TL: Cond	(1,18)	—	—	5.85*		4.18†
Group × Lat × Cond	(1,33)	—	—	3.01†		—
Lat × Cond	(1,33)	—	—	—	5.12*	
Medial: Cond	(1,34)	—	—	—	5.17*	
Ant × Hem × Cond	(1,33)	—	—	—	4.96*	
Frontal: Hemi × Cond	(1,34)	—	—	—	4.85*	
Frontal: Right Hemi: Cond	(1,34)	—	—	—	11.2**	
Central: Hemi × Cond	(1,34)	—	—	—	2.97†	
Lat × Ant × Cond	(2,66)	3.45*	—	—	—	—
Group × Ant × Hemi × Cond	(2,66)	—	3.66*	—	—	—
DLD: Ant × Hemi × Cond	(2,30)	—	4.33*	—	—	—
Group × Lat × Hemi × Cond	(1,66)	—	—	—	5.40*	—
DLD: Lat × Hemi × Cond	(1,15)	—	—	—	4.66*	—
DLD: Medial: Hemi × Cond	(1,15)	—	—	—	2.88†	—
DLD: Lateral: Hemi × Cond	(1,15)	—	—	—	4.31†	
Midline electrodes						
Condition	(1,33)	—	—	—	4.75*	—
Groupe	(1,33)	—	—	—	—	—
Group × Cond	(1,33)	—	—	7.42**	—	7.76**
DLD: Cond	(1,15)	—	—	—	—	
TL: Cond	(1,18)	—	—	7.30*	—	6.92*
Electrodes × Cond	(2,99)	—	—	—	2.40†	—

Note. Only significant results and trends are presented. Cond = Condition; Ant = Anteriority; Lat = Laterality; Hemi = Hemisphere. †: $p < 0.10$, *: $p < 0.05$, **: $p < 0.01$, and ***: $p < 0.001$.

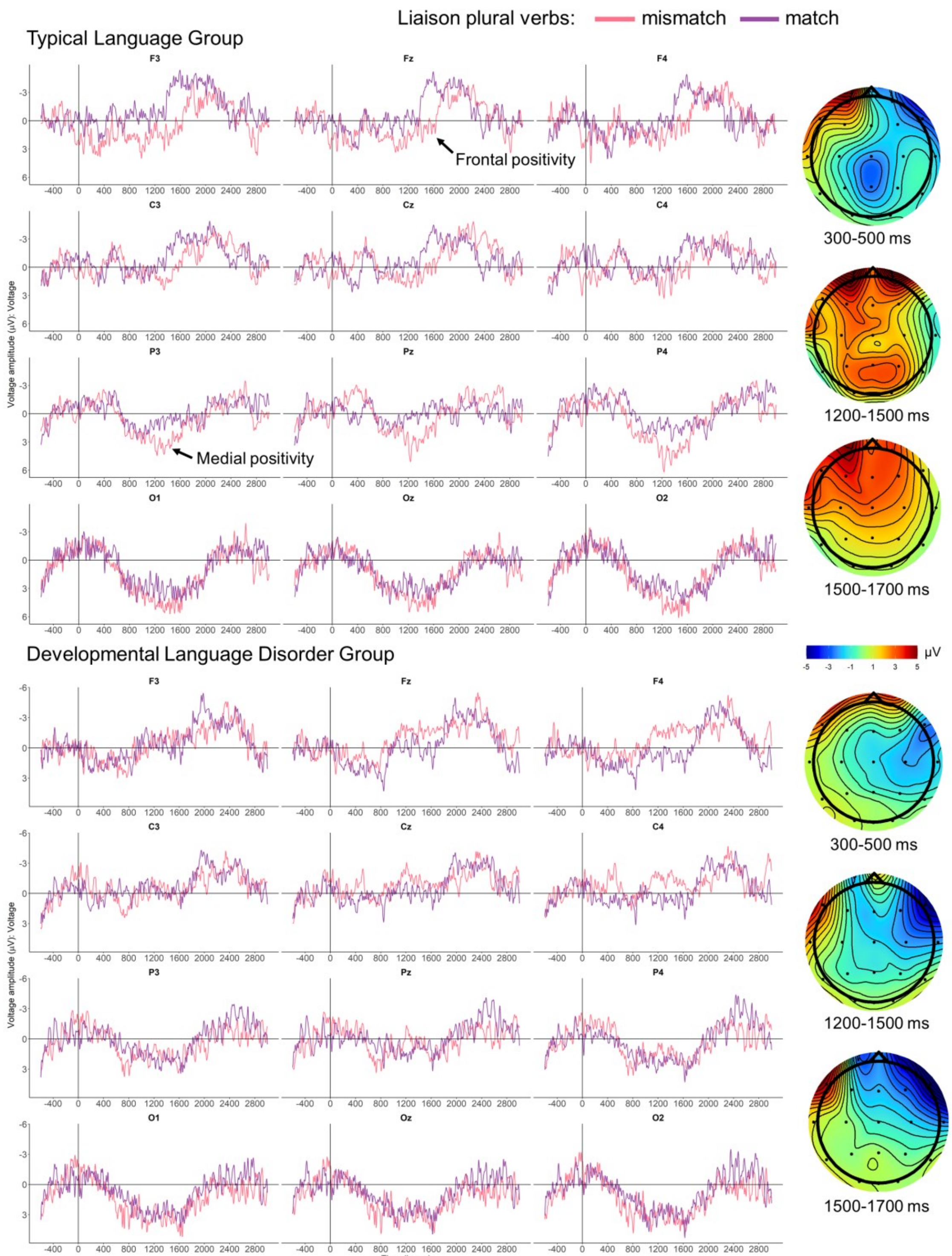
In a nutshell, the CONS verb singular mismatches elicited a significant positivity broadly distributed for the TL group from 1000 to 1400 ms, and a significant positivity in frontal channels of the right hemisphere for all participants from 1400 to 1800 ms, followed by a sentence-final negativity for the TL group only. These results raised the possibility that the two positivities from 1000 to 1800 ms reflected the same ERP component, but with an earlier onset for the TL group (1000 ms) versus the DLD group (1400 ms). We ran additional ANOVAs directly comparing the two time-windows. As expected, Time-Window showed significant interactions with Group and topographical factors, confirming that the P600-like positivity from 1000–1400 ms was elicited in the TL group only, and was a different ERP component from the positivity in the 1400–1800 time-window (see Supplementary Materials Section 4 for the details). As an additional precaution, we ran an ANOVA on both groups separately in order to comfortably reject the possibility that the positivity in the 1400–1800 time-window could have been driven by the DLD group only. If this was the case, we would find a CONDITION effect when running the analyses within the DLD group only. We found no significant effect involving CONDITION when analysing groups separately, which confirmed that this frontal positivity was a shared effect present in both groups.

6.3.2.3.4. ERPs for LIAIS plural conditions

While consonant-final verbs provided number information on the verb-final consonant, liaison number cues were carried by the presence or absence of the liaison phoneme /z/ at pronoun offset. To avoid baseline issues, we calculated ERPs at pronoun onset for both plural and singular conditions, as shown in Figure 6 and 7. Plural LIAIS mismatches elicited a small N400-like negativity between 300–500 ms at Pz and Cz in the TL group (Figure 6), followed by a P600-like positivity in parietal and central electrodes from 1200–1500 ms and a frontal positivity that emerged around 1500 ms. In the DLD group, a right-lateralized fronto-central negativity is found from 1200 to 1700 ms. Based on visual inspection, we selected 300–500, 1200–1500 and 1500–1700 ms time-windows for statistical analyses.

Figure 6

ERP number mismatch effects for liaison plural verbs in neutral contexts.



Note. Grand-average ERPs for the TL group (above) and the DLD group (below) are displayed at midline and eight lateral electrodes, as well as voltage maps illustrating difference waves time-locked to pronoun onset , using a baseline of -600 to 0 ms. Compared to the match condition (purple), plural mismatches (pink) elicited a right-lateralized negativity in lateral channels from 300–500 ms in both groups. In the TL group, this was followed by a P600-like positivity from 1200–1500 ms, significant in medial electrodes, which shifted to a more frontal distribution from 1500–700 ms. The DLD group also exhibited a positivity, but only in the more external channels of the left hemisphere. This effect seems to be driven by horizontal eye movements given its topography and a polarity inversion between both hemispheres, as illustrated on their 1200–1500 ms voltage map.

Global ANOVAs (see Table 11) on the 300–500 ms time-window revealed significant interactions of CONDITION and topographical factors in lateral channels. Decomposition of HEMISPHERE × CONDITION and ANTERIORITY × HEMISPHERE × CONDITION interactions revealed a negativity in the right-hemisphere central channels in both groups. We didn't find any significant effects in the midline electrodes, showing that the apparent N400-like negativity in the TL group didn't reach significance. Between 1200 and 1500 ms, the ANOVA yielded a significant GROUP × LATERALITY × CONDITION interaction which, when decomposed, confirmed the positivity for the TL group in medial lateral electrodes. This positivity did not reach significance at midline channels, when decomposing the GROUP × CONDITION interaction. In order to confirm that the apparent effect in Pz, as seen in the TL voltage map (Figure 6), was significant and part of the same positivity as observed at medial–parietal electrodes, we ran an additional ANOVA for the P3, P4, Pz channels with the factors LATERALITY (Medial vs. Midline) and CONDITION. We found a significant effect of CONDITION, and no significant interaction, supporting our interpretation that the positivity was present in all parietal channels. We also found in the 1200 and 1500 ms time-window an effect shared by both groups as indicated by a significant HEMISPHERE × LATERALITY × CONDITION interaction in the lateral electrodes, which is subsumed by a significant effect of CONDITION in the more lateral channels of the left hemisphere. As illustrated on the 1200–1500 ms voltage map (Figure 6), in the DLD group a positivity was elicited in left lateral electrodes (i.e., F7, T7, P7), which explains this shared effect despite the fact that their voltage map is mainly negative. Note that the channels carrying this positivity in the DLD group are the ones that tend to be influenced by muscle movements and horizontal eye movement artifacts, which is what we believe is behind this pattern. This assumption seems to be supported by the distribution of effects on the scalp, where we observe a polarity inversion of this difference between left-anterior and right-anterior electrodes—especially F7-T7 and F8-T8, which is typical of horizonal eye-movement artefacts. Since we targeted only blinks (i.e., vertical eye movements) with the ICA artefact rejection procedure, it's possible that horizontal eye movements are contaminating these external channels. This positivity in left lateral electrodes continued for both groups in the 1500–1700 ms time-window, as supported by significant interactions involving ANTERIORITY, LATERALITY, HEMISPHERE and CONDITION. The TL positivity with a frontal distribution was confirmed in both lateral and midline channels from 1500 to 1700 ms, as supported by ANTERIORITY × GROUP × CONDITION and ELECTRODE × GROUP × CONDITION interactions and their decompositions.

Table 11

Global ANOVAs for liaison plural verbs at verb onset at time-windows of interest

		Negativity	Positivity	
	df	300-500	1200-1500	1500-1700
LATERAL ELECTRODES				
CONDITION	(1,33)	—	—	—
GROUP		—	—	—
HEMI × COND	(1,33)	7.33**	5.75*	6.22*
HEMI R: COND	(1,34)	4.57*	—	—
HEMI L: COND	(1,34)	—	4.24*	5.13*
ANT × HEMI × COND	(2,66)	5.85*	8.13**	7.37**
FRONTAL: HEMI × COND	(1,34)	8.41**	5.93*	—
FRONTAL: RIGHT HEMI: COND	(1,34)	3.49†	—	—
FRONTAL: LEFT HEMI: COND	(1,34)	—	4.05†	—
CENTRAL: HEMI × COND	(1,34)	5.35*	10.93**	—
CENTRAL: LEFT HEMI: COND	(1,34)	—	4.11†	—
CENTRAL: RIGHT HEMI: COND	(1,34)	4.35*	—	—
LEFT HEMI: ANT × COND	(1,34)	—	—	4.30*
LEFT HEMI: FRONTAL: COND	(1,34)	—	—	7.39**
LEFT HEMI: CENTRAL: COND	(1,34)	—	—	4.49*
ANT × LAT × COND	(2,66)	5.22*	—	—
HEMI × LAT × COND	(1,33)	—	4.45*	10.73**
LATERAL: HEMI X COND	(1,34)	—	5.79*	7.61**
LATERAL: LEFT HEMI: COND	(1,34)	—	4.67*	8.87**
HEMI × ANT × LAT × COND	(2,66)	—	—	8.73**
LEFT HEMI: ANT × LAT × COND	(2,68)	—	—	3.82*
LEFT HEMI: FRONTAL × LAT × COND	(1,34)	—	—	4.57*
LEFT HEMI: FRONTAL: LATERAL: COND	(1,34)	—	—	12.00**
GROUP × LAT × COND	(1,33)	—	7.70**	—
TL: LAT × COND	(1,18)	—	6.22*	—
TL: MEDIAL : COND	(1,18)	—	5.21*	—
ANT × GROUP × COND	(2,66)	—	—	5.16*
FRONTAL: GROUP × COND	(1,34)	—	—	5.40*
FRONTAL: TL: COND	(1,18)	—	—	5.88*
CENTRAL: GROUP × COND	(1,34)	—	—	3.11†
GROUP × ANT × HEMI × LAT × COND	(2, 66)	—	4.02*	7.75**
DLD: ANT × HEMI × LAT × COND	(2,30)	—	8,17**	—
DLD: FRONTAL: HEMI × LAT × COND	(1,15)	—	7.61**	—
DLD: FRONTAL: LEFT HEMI: LAT × COND	(1,15)	—	5.28*	—
DLD: FRONTAL: LEFT HEMI: LAT: COND			3.15†	—
DLD: CENTRAL: HEMI × LAT × COND	(1,15)	—	8.83**	—
DLD: CENTRAL: LEFT HEMI: LAT × COND	(1,15)	—	5.28*	—
DLD: CENTRAL: LEFT HEMI: LAT COND	(1,15)	—	4.39†	—
LEFT HEMI: ANT × LAT × GROUP × COND	(2,66)	—	—	5.86*
LEFT HEMI: FRONTAL: LAT × GROUP × COND	(1,33)	—	—	10.09**
LEFT HEMI: FRONTAL: MEDIAL: GROUP × COND	(1,33)	—	—	4.06†
MIDLINE ELECTRODES				
CONDITION	(1,33)	—	—	—
GROUP	(1,33)	—	—	—
GROUP × COND	(1,33)	—	4.69*	—
TL: COND	(1,18)	—	4.35†	—
ELECTRODE × GROUP × COND	(3,99)	—	—	3.63*
FZ: GROUP × COND	(1,33)	—	—	4.69*
FZ: TL: COND	(1,18)	—	—	4.60*

Note. Only significant results and trends are presented. Cond = Condition; Ant = Anteriority; Lat = Laterality; Hemi = Hemisphere. †: $p < 0.10$, *: $p < 0.05$, **: $p < 0.01$, and ***: $p < 0.001$.

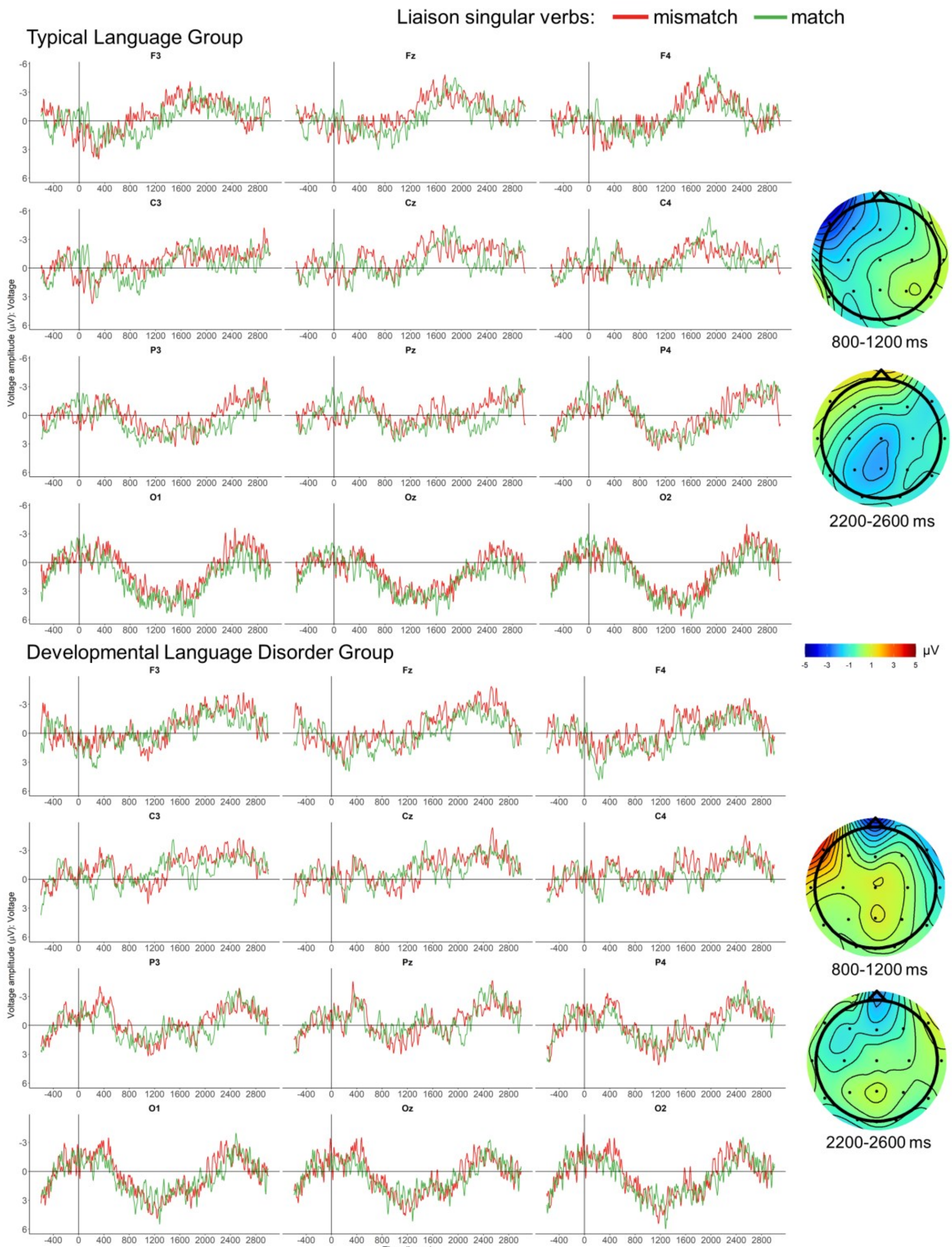
To sum up, in both groups plural LIAIS verb mismatches first elicited a right-lateralized negativity in central channels (i.e., C4 and T8) from 300 to 500 ms. Focusing on C4 in Figure 6, we can see that in the TL group this effect does not look like a typical ERP negativity, but rather noise, considering that the difference between conditions moves from positive to negative between 300–500 ms. In the DLD group, this right-lateralized negativity seemed to be part of the same polarity inversion pattern for temporal electrodes that we found in the other time-windows (1200–1500, 1500–1700) which is typical of artefacts and not of cognitive processes. Because of these issues, we refrain from interpreting these differences. From 1200 to 1700 ms, we found in the TL group a small P600-like positivity that was significant only in medial channels (1200–1500 ms) and evolved with a frontal distribution (1500–1700 ms).

6.3.2.3.5. ERPs for LIAIS singular conditions

Figure 7 depicts number mismatches for singular LIAIS verbs. We observe an apparent LAN-like component at F3 in the TL group from 800 to 1200ms and a small P600-like positivity in the DLD group. Starting at 2200 ms, a sentence-final negativity can be seen in the central and parietal midline channels of the TL group only. Based on these observations, we selected time-windows from 800–1200 ms and 2200–2600 ms for analyses.

Figure 7

ERP number mismatch effects for liaison singular verbs in neutral contexts.



Note. Grand-average ERPs for the TL group (above) and the DLD group (below) are displayed at midline and eight lateral electrodes, as well as voltage maps illustrating difference waves, time-locked to pronoun onset (vertical bar) using a baseline of -600 to 0 ms. Compared to the match condition (green), plural mismatches (red) did not elicit any consistent ERP pattern in either group.

Results are summarized in Table 12. Global ANOVAs for the first time-window revealed effects in the TL group only, as supported by the significant interactions involving topographical factors, GROUP and CONDITION. This LAN-like negativity was significant only in the more lateral electrodes, namely F7, and was driven by one participant. Since F7 is more prone to be influenced by eye movements, and that singular LIAIS verb conditions elicited no effect in adults (Courteau et al., 2019) we refrain from interpreting this negativity only supported by this one channel. Between 2200–2600 ms, the only effect we found was a trend in midline electrodes for a GROUP × ELECTRODE × CONDITION interaction, indicating that the apparent TL-group sentence-final negativity did not reach significance.

Table 12

Global ANOVAs for liaison singular verbs at verb onset at time-windows of interest

	df	800–1200	2200–2600
LATERAL ELECTRODES			
CONDITION	(1,33)	—	—
GROUP		—	—
GROUP × HEMI × COND	(1,33)	5.67*	—
TL : LEFT HEMI X COND	(1,18)	5.35*	—
GROUP × ANT × HEM × COND	(2,66)	4.86*	—
FRONTAL: GROUP × HEM × COND	(1,33)	6.10*	—
FRONTAL: TL: HEM × COND	(1,33)	7.76*	—
FRONTAL: TL: LEFT HEMI: COND	(1,18)	6.56*	—
GROUP × LAT × HEM × COND	(1,33)	4.95*	—
LATERAL: GROUP × HEM × COND	(1,33)	6.06*	—
LATERAL: TL: HEM × COND	(1,18)	4.30 †	—
GROUP × LAT × ANT × HEM × COND	(2,66)	6.00*	—
TL: LAT × ANT × HEM × COND	(2,36)	4.30*	—
TL: LATERAL: ANT × HEM × COND	(2,36)	5.39*	—
TL: LATERAL: FRONTAL: HEMI × COND	(1,18)	6.06*	—
TL: LATERAL: FRONTAL: LEFT HEMI: COND	(1,18)	13.37*	—
MIDLINE ELECTRODES			
CONDITION	(1,33)	—	—
GROUP	(1,33)	—	—
GROUP × ELECTRODE × COND	(1,99)	—	2.80 †

Note. Only significant results and trends are presented. Cond = Condition; Ant = Anteriority; Lat = Laterality; Hemi = Hemisphere. †: $p < 0.10$, *: $p < 0.05$, **: $p < 0.01$, and ***: $p < 0.001$.

Overall, our analyses did not point to any consistent ERP pattern for singular LIAIS verb mismatches in both groups. This is not surprising considering that LIAIS singular verb mismatches did not induce significant effects in the adult's group as well (see the discussion section and Courteau et al., 2019, for details on why). For the TL group, we found an effect at F7 that we judged to be unreliable.

6.4. Discussion

The current study aimed to compare the predictions of two accounts for development language disorder, the PDH and the GSH. We compared number and lexico-semantic processing in (pre-)teens with and without DLD while they listened to grammatical sentences and looked at pictures that matched or not their lexico-semantic or morphosyntactic properties. Processing was described at a behavioural level with grammaticality-judgement accuracy and response times, and at a neurocognitive level using ERP components types, distribution and timing. Depending on the conditions, results showed similarities and differences, both qualitative and quantitative, between groups. We will first discuss behavioural results. Next, in relation to previous ERP studies and from a developmental perspective (Table 1), we will present ERPs elicited by lexical-semantic and morphosyntactic conditions. Finally, we will discuss our results from the perspectives of the PDH and the GSH.

6.4.1. Behavioural performance

Accuracy of acceptability judgements first revealed that, while the TL group clearly outperformed the DLD group on morphosyntactic mismatches, as supported by large effect sizes, both groups performed similarly on sentences containing lexico-semantic mismatches. Studies that have compared lexical-semantic and morphosyntactic grammaticality judgement skills have generally found, using ungrammatical sentences, that TL groups perform better than DLD ones across all conditions when investigating children, as Pawlowska et al. 2014, and teenagers, as Haebig et al. 2017. Even though our task was possibly more cognitively demanding for our participants than only listening to ungrammatical sentences, as they had to integrate visual and auditory information, the nature of our task could have helped participants with DLD. Indeed, the picture provided a visual support that was present during the whole time they were hearing the sentence. This could have enhanced their performance and helped them achieve their full potential when they had to make the judgements at the end of the trial, which translates to similar performances in both groups on the lexico-semantic conditions. One interpretation for this good performance lies in the fact that lexico-semantic anomalies lasted longer in time and were supported by more cues than morphosyntactic ones. The mismatches in the former conditions were supported by multiple words that did not match the picture, and in the latter with the addition or absence of one or occasionally two phonemes. The fact that lexico-semantic mismatches were supported by multiple words could have helped participants with DLD, as some are known to have phonological working memory deficits (Archibald & Joanisse, 2009). This suggests that the presence of multiple cues can help DLD participants perform better on acceptability judgements, although this was not the main question we addressed in our experiment.

More cues also enhanced DLD participants performance within the morphosyntactic conditions, on plural conditions (commission) and sentences with NP context. Interestingly, this effect does not always apply to the TL group. Indeed, in the DLD group, plural conditions, with commissions errors, were slightly better identified than singular ones, i.e., omissions. This was not the case for the TL group who performed similarly on both singular and plural sentences. Sentences with subject NPs, including two number cues, were better identified by the DLD group than sentences with neutral contexts and only one cue, as supported by a moderate effect size. However, in the TL group, the presence of subject NPs only slightly improved their performance, as indicated by a small effect size. This finding could be in line with working memory-based accounts of DLD (e.g., Montgomery et al., 2016), which posits that individuals with DLD have memory storage deficits that affect their ability to understand longer and more complex sentences. Indeed, our results suggest that repeating the number cues had a larger positive impact the DLD group performance than for the TL group, which is predicted by working memory-based accounts of DLD. However, a main effect of verb type suggests that more cues and working memory only could not explain the performance of DLD participants in all conditions. We found that LIAIS verbs were slightly better judged than CONS ones across the groups, even though CONS as overall more phonemic cues than LIAIS. This imply that regular subject-verb agreement (liaison) was better identified than irregular consonant-final verbs by our participants. Note that this result should be taken with caution as it was characterised only by a small effect size.

Regarding response times, no significant effect involving group was found: both groups took the same time to perform acceptability judgements. This is surprising given the multiple studies that found slower response times for children and teenagers with DLD when compared to TL groups (Miller Carol A. et al., 2006; Quémart & Maillart, 2016; Royle et al., 2002). We think this absence of effect could be related to our experimental design. Following sentence offset, participants had to wait 1000 ms before the response prompt appeared on the screen. This was done to avoid contamination of muscle-related components in the ERPs. We think that this time buffer makes our experiment less sensitive to potential timing differences between our groups.

6.4.2. Event-related potentials

6.4.2.1. Lexico-semantic conditions

Lexico-semantic mismatches on verbs elicited similar ERP patterns in both groups, which were primarily two components: an N400 followed by a sentence-final negativity, with only quantitative differences between groups. The topography of the N400 was similar in both groups, but in the DLD group the onset of the component was delayed, and its offset was earlier. While some studies find comparable N400 onsets latencies across participants with and without DLD, as Haebig et al. (2017) and Kornilov et al. (2015), others have revealed delays for the ones with DLD (Neville et al., 1993, Pijnacker et al., 2017: see the introduction section for an overview of these studies). How could we explain these disparities between these results? A first explanation could be the age of the participants, where only younger participants with DLD would show N400 onset delays when compared to their peers with TL. Indeed, studies reporting delays had younger participants, as Neville et al. (1993) with 9-year-olds, and Pijnacker et al. (2017) with 5-year-olds, whereas studies that found similar onsets, such as Haebig et al. (2017) and Kornilov et al. (2015), tested groups with participants over 10 years old. This pattern would suggest longer maturation of lexical-semantic processes in children with DLD when compared to TL children, but DLDs would catch up by late childhood as delays are not observed after 10 years old. Consistently, delays in N400 onset are a characteristic of children with TL when compared to adults (Cummings et al., 2008; Juottonen et al., 1996). However, our participants with DLD were 14-year-old adolescents, so age alone could not explain the N400 onset delays. A second explanation could be tasks' modality, where visual or bimodal information processing would elicit onset delays in N400s of DLD participants. Nevertheless, Kornilov et al. (2015) used a word-picture paradigm mismatches and found similar onset across groups with and without DLD. In our opinion, experimental task complexity is the best candidate to account for N400 onset delays in participants with DLD. The complexity of a task is increased when it involves sentence processing, in contrast to isolated word processing, and when it uses the visual or bimodal modality as opposed to the auditory modality only. Studies that didn't find delays used either an isolated word-picture matching task, as in Kornilov et al. (2015), or the auditory-only modality to investigate sentence processing, as in Haebig et al. 2017. In comparison, studies that found delays, including ours, used more complex tasks, as they investigated sentence processing using visual or bimodal modality. Neville et al. (1993) and Pijnacker et al. (2017) investigated sentence processing with a final semantically incongruent noun. In the former study, participants read sentences, and in the latter participants listened to sentences while looking at silent unrelated videos. Our task required that the lexical representation of the verb be activated by the picture. When participants heard the verb, they had to process the (mis)match between phonologically- and visually-activated semantic representation. If, as our results suggests, N400 onset delays found in DLD participants during lexico-semantic processing are related to task complexity and not age, future studies should use complex tasks and target older participants as young adults or adults with DLD to see if delays are still present. Another quantitative difference between our groups was the offset of the N400. It lasted until 950 ms in the TL group but ended earlier in the DLD one, at 750 ms. In adults, our lexico-semantic conditions elicited an N400 that was long-lasting. Courteau et al. (2019) proposed multiple interpretations for this longer duration, including the fact that audio-visual mismatches were not only on the verb but also on the following prepositional phrase (see Table 3). The earlier offset of the N400 of participants with DLD suggests that they probably didn't process the mismatches following the verbs, at least not in the same way as the TL group did.

The pattern of an N400 followed by a sentence-final negativity is reminiscent of adults in the same condition (Courteau et al., 2019). This effect was elicited in both the TL and DLD groups from 2000 to 2650 ms. A negativity following the offset of a sentence containing anomalies has been argued to reflect additional processing load involved in reconsidering the mismatch and integrating the whole sentence (Osterhout & Mobley, 1995), and to be modulated by task (Stowe et al., 2018). Here both groups elicited this effect, showing that the DLD group were able to reconsider the mismatch and integrate the whole sentence in this condition. This is not surprising as both groups processed the lexico-semantic mismatches, as exhibited by similar N400s with only quantitative differences, and the same performance on the accuracy judgement task.

6.4.2.2. Morphosyntactic conditions

The morphosyntactic mismatches elicited N400s, P600s or no effect, depending on conditions. No adult-like biphasic ERP pattern was found in either group. Indeed, adults in the same experiment (Courteau et al., 2019) elicited negativities followed by positivities in all conditions except for singular LIAIS verbs, where no effect was found. This suggests that our participants with TL, aged 12 years old on average, are still maturing and learning how to process number agreement. This interpretation is coherent with previous literature, as Cantiani et al. (2015) and Haebig et al. (2017) found only P600s in their (pre-)adolescents with TL, and not an adult-like biphasic pattern. However, also based on previous literature (see Table 1), we expected our participants to exhibit slight variability across conditions, namely 1) more adult-like ERP patterns when processing morphosyntactic markers acquired earlier, as regular number agreement, and 2) more immature patterns when processing late-acquired structure, including irregular agreement. Results are coherent with our expectations within our conditions for regular agreement. We will discuss how sociolinguistic factors, pragmatic interpretations, and the particularities of the French language could also have influenced our results. We will first discuss the results of sentence onset and LIAIS verb mismatches, reflecting regular agreement, followed by CONS verbs, characterised by irregular agreement.

For regular agreement conditions, markers acquired earlier elicited ERP patterns corresponding to a higher development level of morphosyntactic processing (see Steinhauer et al., 2009 and Table 1). Recall that determiner agreement number marking in NPs, as in our sentence onset subject context conditions, are understood and produced by French-speaking children by 1;8 years old (Valois et al., 2009), slightly earlier than pronoun-verb liaison which is processed by the age of 3 years-old (Legendre et al., 2014). In the TL group, both sentence onset subject context and LIAIS verb number agreement mismatches elicited positivities. However, the former revealed a large P600, which corresponded to an intermediate level of morphosyntactic processing, and the latter a small delayed P600, on the verge of significance in the midline channels, corresponding to a low to intermediate level. In contrast, the DLD group exhibited a small delayed P600 for singular subject context mismatches and no effect for LIAIS verb mismatches, also showing a more precocious level of development for earlier acquired morphosyntactic cues. It might be surprising that cues acquired in such proximity during early childhood point towards different levels of morphosyntactic processing for our (pre-)teen participants. The consistency with which these agreement markers are produced in French oral language may also have played a role. On the one hand, determiner agreement number in NPs is obligatory and consistently produced (Valois et al., 2009) while on the other hand, liaison on verbs is not productively used in French-speaking adults in France (less than 20 %, Legendre et al., 2014) and in Quebec French, *ils* 'they' before a vowel-initial verb can be realised without a plural /z/ (e.g., /ijɔ̃/ for *ils ont* 'they are'; Durand & Lyche, 2008; that is, it is neutralized and thus not always perceptible, depending on dialect.

At sentence onset, singular mismatches elicited significant condition effects in both groups, whereas no consistent effect was found in the plural conditions. To understand this asymmetry, note that mismatches at sentence onset with subject contexts could have promoted processing based on pragmatics and logical truth values, instead of purely morphosyntax (Barbet & Thierry, 2016; Courteau et al., 2019). Participants were looking at a picture involving one or two subjects when they heard the sentence onset, which was either a singular or plural determiner. For example, in the singular condition, the image of two girls buying candies accompanied by the sentence *La fille achète* 'The girl buys' could be interpreted as logically true, as for instance the sentence *Some triangles have three edges*, but pragmatically odd, since there was only one girl in the picture. Inversely, in the plural condition, an image of one girl with the sentence *Les filles achètent* ‘The girls buy’ is logically false, but pragmatically relevant as it could be interpreted as a generalisation: in French, one can state that “all girls like” with the definitive plural form which was used. Our results indicated that both groups rejected the pragmatically odd mismatches, as illustrated by the elicited positivities, but they did not matter the pragmatically appropriate pairs, i.e., the generalisation, as plural mismatches elicited no effect. Interestingly, previous studies on children’s comprehension of pragmatic and logical meaning have shown that children have a preference for the logical interpretation (Katsos et al., 2011; Noveck, 2001). Adults, however, show a preference for the pragmatic interpretation (Papafragou & Musolino, 2003) or they will differ in their bias toward logical vs. pragmatic interpretations (Barbet & Thierry, 2016; Noveck, 2001). Our ERP effects were clearly aligned with the pragmatic interpretation, pointing to behaviour similar as adults in teens, which is not surprising considering their ages. However, our adults' ERP pattern (Courteau et al., 2019) did not differ between singular and plural mismatches. This could indicate that the teenager participants behave only partly like adults, aligning themselves to the pragmatic interpretation, but, unlike adults, they are more categorical and reject the logical interpretation. Future studies should target adolescents to see at what ages we should expect adult-like behaviour. In the singular condition, where we found an effect, both groups displayed a positivity. The TL group displayed a large P600 that was present from 600 to 1200 ms, and the DLD group a delayed P600 that increased from 600 to 800 ms and peaked between 800-1000 ms. How can we interpret this delay? As seen in Table 1, a P600 and a delayed P600 are associated with two different developmental levels of processing, and this delay suggests different underlying processes used by our groups. This difference in processing could be related to the fact DLD are know to have impairments not only in morphosyntax but also in pragmatics (Bishop et al., 2017). This is consistent with Katsos et al. (2011) who observed difference between groups of children with and without DLD on pragmatic and logical interpretations. Overall, our results suggest that even when processing early acquired markers such as number agreement in NPs, differences are still found during adolescence between youth with and without DLD.

Processing of liaison verbs in the singular conditions elicited no effect in either group. This is not surprising since the adult group in the same experiment did not process the error in this particular condition either. Courteau et al. (2019) proposed as the most plausible explanation for this lack of effect sociolinguistic factors, where in Quebec French, verbs with a vowel onset in the plural form can be produced without the liaison (Durand & Lyche, 2008; see Courteau et al., 2019 for a detailed explanation). In contrast to singular mismatches, plural mismatches did elicit a reliable biphasic ERP effect in adults, i.e., when hearing the plural liaison while seeing a picture of just one subject (Courteau et al., 2019). Similarly, the TL group in our present study exhibited a small P600 in response to plural liaison mismatches, whereas the DLD group showed no significant component. The most credible explanation for these ERP patterns is again the fact that liaison on plural verbs is not frequent and nor obligatory in Quebec French oral language, despite the fact that it is a regular rule in the standard French. It is known that children with and without DLD are influenced by the frequency of the input when learning language (e.g., Leonard et al., 2015). One might think that liaison itself could be a specific domain of deficit in DLD, but Chevrot et al. (2007) demonstrated that children with DLD were able to detect obligatory liaison with the phoneme /z/ between determiners and nouns in French (e.g., *deux‿ours* ‘two bears’) the same way as children with TL during grammaticality judgement.

Turning to the processing of late-acquired structure, as our irregular CONS verbs, singular and plural conditions elicited ERP effects in both groups. Singular mismatches elicited a P600 and a sentence-final negativity in the TL group, and a late frontal-positivity for the DLD group. Singular mismatches on CONS verbs are arguably the most subtle morphosyntactic condition. As opposed to both sentence onset and LIAIS verb conditions, the number cue is at the end of the verb, and the cue is an omission of the verb-final consonant, which is not very salient. This could explain why the TL group showed processes of sentence reanalysis, as evidenced by a P600 followed by a sentence-final negativity, whereas the DLD group presented only a late frontal positivity. This frontal positivity could be interpreted as a late P3a which indexes domain-general engagement of attentional processes (Courteau et al., 2018; Van Petten & Luka, 2012), thus indicating that participants with DLD noticed that there was something unusual with the stimuli but didn't engage in sentence-repair processes. Plural CONS verb mismatches, on the other hand, elicited an N400 in both groups, which could be interpreted initially as a rather low level of morphosyntactic processing by our participants (see Steinhauer et al., 2009 and Table 1). This low level could indicate that our participants were still consolidating the plural form of these verbs. During an induced oral production task using the same verbs as in our experiment, half of the participants in the TL group were still making between 5 and 10 % of errors, and the DLD group made more than 20 % of error on average (Courteau et al., in revision). However, we suspect that this would not be the only reason for the appearance of this N400 in reaction to irregular verb mismatches. Indeed, even if participants in the TL group were still consolidating production of CONS plural verb, it would be surprising that these elicited more immature morphosyntactic processing than plural LIAIS verbs, which are produced in French oral language less than 20 % of the time by adults (Legendre et al., 2014). Rather than signalling a very low level of morphosyntactic processing, another plausible explanation for these N400s is that they may represent the expected reflection of lexical access for irregular plural verb forms. Adult participants produced a broadly distributed N400 followed by a P600 in this condition (Courteau et al., 2019). Both TL and DLD groups could have detected the anomaly and correctly engaged lexico-semantic processes, as illustrated by the N400, but didn't repair it, in contrast to adults who deployed their repair processes as indexed by the P600. Clahsen et al. (2007) found, in response to plural noun overregularization errors in German, a biphasic pattern starting at age 8. The absence of adult-like pattern in our participants could be attributed first to the fact we used verbs, which have a more complex lexical representation than nouns, and second that our plural stimuli, characterised by nine different possible consonantal verb endings, were more irregular than the ones used in Clahsen et al. (2007)'s study.

6.4.3. Two accounts for DLD and their predictions

Overall, our behavioural and neurocognitive results better fit the PDH framework (Ullman & Pierpont, 2005; Ullman et al., 2020) than the GSH (Kail, 1994), with some caveats. The PDH predicts that conditions that are underpinned by the declarative memory system, such as the lexical-semantic one, would induce similar results across groups. This is what we observed at the behavioural level, with similar accuracy judgements *A*-scores, as well as at the neurocognitive level, where both groups elicited an N400 followed by a sentence-final negativity. Quantitative differences were found between groups: DLD group's N400 onset was 100 ms later than the TL one and had an earlier offset. One explanation for this delay could be that our tasks tapped into lexical retrieval processes, which are expected to be underpinned by the procedural memory, according to the PDH. However, our task involved lexical comprehension and participants did not have to provide a rapid acceptability judgement, which should reduce the implication of lexical retrieval processes. Furthermore, on the basis of the PDH, these quantitative differences, such as component latencies, do not rule out that the two groups used similar processes. Given the previous literature and the current results, the most plausible explanation for this onset delay is task complexity related to its visuo-auditory modality. Furthermore, the ERP patterns found in this condition are similar to those found in adults (Courteau et al., 2019), confirming that adolescents with DLD can deploy lexical-semantic strategies similar to those in mature adults for lexical-semantic processing. Based on the PDH, it was also expected that processing of irregular verbs agreement would recruit, at least at the word level, lexical-semantic processes. This was observed in our groups on the CONS plural verb mismatches where both participant groups elicited N400s. Regarding rule-governed morphosyntactic processing at both the word-form and syntactic levels, LIAIS plural verb results aligned with the PDH predictions, where we found qualitative differences between groups: no effect for the DLD group and P600s for the TL group. This lack of effect for the DLD group in this condition is coherent with the PDH, as it predicts that the declarative system will not be able to easily compensate for low-frequency ruled-governed processes (Ullman et al., 2020), which is the case for number agreement liaison on verbs in French. Mismatches on sentence-initial subject context elicited P600s in both groups, with only quantitative differences: a large P600 in the TL group and a delayed and shorter one in the DLD group. The PDH would have predicted qualitative differences, as the TL group should have used processes subserved by the procedural system and the DLD group declarative ones. However, P600s are not presumed to depend on procedural memory (Ullman 2020, p. 147), in contrast to anterior negativities and LANs, therefore our results are still theoretically aligned on the PDH.

The GSH predictions were that, while both groups should use similar processing approaches for lexical-semantic and morphosyntactic conditions, we should observe processing speed deficits in the DLD group, instantiated by longer response times and delays in ERP components compared to the TL group. We found no indicator of deficits in processing speed at the behavioural level, as accuracy judgment reaction times did not differ between groups. This lack of difference is most likely related to our experimental design, which included a buffer of 1000 ms between sentence offset and accuracy judgement. At the neurocognitive level, we found processing delays in two conditions: on the onsets of the N400 for lexico-semantic mismatches and of the P600 for sentence onset subject context mismatches. However, we did not observe a processing limitation in other conditions, but rather qualitative differences or no timing difference at all. The GSH is based on the premise that our participants will use the same processes to complete our tasks. This is not what we observed. Instead, our study provides evidence that (pre-)teenagers with and without DLD differ in the processes they engage in to understand morphosyntax in comparison to lexical-semantic information. This was seen at the behavioural level, with different accuracy judgement performance when considering only morphosyntax, as well at the neurocognitive level through ERP components.

Although our results better support the PDH, we suggest that, in the study of DLD, the assessment of processing speed through task response times remains relevant and should be more systematically addressed in research. Instead of considering teens with DLD as having deficient processing speed i.e., the cognitive function as defined by Kail and Salthouse (1994), we propose that delays observed during task completion should be considered a consequence of impaired processes underlying language. This would be coherent with research on developmental dyslexia, where slower reading speed is observed as a result of impaired reading processes (Ziegler et al., 2008)).

The primary limitation of this study was that the participants with TL were slightly younger and had a wider age range than those with DLD. With older participants with TL, we could have observed in the morphosyntactic conditions more mature patterns for them and thus greater differences between our groups. As mentioned above, future studies should target older adolescents and young adults. Our results and the previous literature indicate that adolescents are still consolidating morphosyntactic processes and do not, overall, show the same ERP patterns as adults. It would be useful and important to study several consecutive adolescent age groups from a developmental point of view, in order to understand the evolution of ERP patterns as they converge to adult ones.

5. Conclusion

Our study provides evidence supporting the PDH, and by the same token that lexico-semantics is a relative strength in teenagers with DLD, in comparison to morphosyntax, when processing linguistic information at the sentence level. It is important to emphasise that this relative strength does not indicate that in different lexico-semantic tasks or real language situations, youth with DLD will not demonstrate linguistic deficits. Rather, this suggests that morphosyntactic processing is comparatively deficient at the behavioural and neurocognitive levels in DLD as compared to the lexico-semantic one.

The current results suggest that ERP patterns elicited during sentence processing in French are, at least in the TL group, aligned with the concept of morphosyntactic development and consolidation during adolescence. Our findings suggest that morphosyntactic markers acquired earlier will induce more mature ERP patterns in (pre-)teens with TL. However, interpretation of results should take into account sociolinguistic factors as well as adults' results in the same experiments to establish what might be expected from expert mature processing. Overall, the ERP technique combined with grammatical sentences is a reliable approach to study similarities and differences between teenagers with and without DLD on neurocognitive processes underlying language comprehension.

6.5. References

American Psychiatric Association. (2013). Neurodevelopmental Disorders. In *Diagnostic and statistical manual of mental disorders DSM-5.* (5th ed., pp. 31–86). American Psychiatric Publishing. https://doi.org/10.1176/appi.books.9781585624836.jb01

Archibald, L. M. D., & Joanisse, M. F. (2009). On the Sensitivity and Specificity of Nonword Repetition and Sentence Recall to Language and Memory Impairments in Children. *Journal of Speech, Language, and Hearing Research*, *52*(4), 899–914. https://doi.org/10.1044/1092-4388(2009/08-0099)

Auger, J. (1995). Les clitiques pronominaux en français parlé informel: Une approche morphologique. *Revue Québécoise de Linguistique*, *24*(1), 21–60.

Barbet, C., & Thierry, G. (2016). Some alternatives? Event-related potential investigation of literal and pragmatic interpretations of some presented in isolation. *Frontiers in Psychology*, *7*, 1479. https://doi.org/10.3389/fpsyg.2016.01479

Bishop, D. V. M., Snowling, M. J., Thompson, P. A., & Greenhalgh, T. (2017). Phase 2 of CATALISE: A multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *Journal of Child Psychology and Psychiatry*, *58*(10), 1068–1080. https://doi.org/10.1111/jcpp.12721

Boersma, P., & Weenink, D. (2017). *Praat: Doing phonetics by computer* (Version 6.0.60) [Computer software]. http://www.praat.org/

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2019). Toward a Neurobiologically Plausible Model of Language-Related, Negative Event-Related Potentials. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.00298

Brucher, A., Courteau, E., Steinhauer, K., & Royle, P. (2021, March 15). *Gender-agreement errors on adjectives and determiners elicit different ERP patterns in French*. Words in the World (WOW) International Conference. https://doi.org/10.13140/RG.2.2.10295.65440

Brunetti, R., Del Gatto, C., & Delogu, F. (2014). eCorsi: Implementation and testing of the Corsi block-tapping task for digital tablets. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00939

Brunner, E., & Munzel, U. (2000). The Nonparametric Behrens-Fisher Problem: Asymptotic Theory and a Small-Sample Approximation. *Biometrical Journal*, *42*(1), 17–25. https://doi.org/10.1002/(SICI)1521-4036(200001)42:1<17::AID-BIMJ17>3.0.CO;2-U

Caffarra, S., Mendoza, M., & Davidson, D. (2019). Is the LAN effect in morphosyntactic processing an ERP artifact? *Brain and Language*, *191*, 9–16. https://doi.org/10.1016/j.bandl.2019.01.003

Cantiani, C., Lorusso, M. L., Perego, P., Molteni, M., & Guasti, M. T. (2013). Event-related potentials reveal anomalous morphosyntactic processing in developmental dyslexia. *Applied Psycholinguistics*, *34*(06), 1135–1162.

Cantiani, C., Lorusso, M. L., Perego, P., Molteni, M., & Guasti, M. T. (2015). Developmental dyslexia with and without language impairment: ERPs reveal qualitative differences in morphosyntactic processing. *Developmental Neuropsychology*, *40*(5), 291–312. https://doi.org/10.1080/87565641.2015.1072536

Chaumon, M., Bishop, D. V. M., & Busch, N. A. (2015). A practical guide to the selection of independent components of the electroencephalogram for artifact correction. *Journal of Neuroscience Methods*, *250*, 47–63. https://doi.org/10.1016/j.jneumeth.2015.02.025

Chevrot, J.-P., Nardy, A., Barbu, S., & Fayol, M. (2007). Production et jugement des liaisons obligatoires chez des enfants tout-venant et des enfants atteints de troubles du langage: Décalages développementaux et différences interindividuelles. *Rééducation Orthophonique*, *229*, 199–220.

Clahsen, H., Lück, M., & Hahne, A. (2007). How children process over-regularizations: Evidence from event-related brain potentials. *Journal of Child Language*, *34*, 601–622. https://doi.org/10.1017/S0305000907008082

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed). L. Erlbaum Associates.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155.

Conti-Ramsden, G., Botting, N., & Faragher, B. (2001). Psycholinguistic markers for specific language impairment (SLI). *Journal of Child Psychology and Psychiatry*, *42*(6), 741–748. https://doi.org/10.1111/1469-7610.00770

Corsi, P. M. (1972). *Human memory and the medial temporal region of the brain (unpublished doctoral thesis)*. McGill University, Montreal.

Courteau, É., Fromont, L., Royle, P., & Steinhauer, K. (2018, March 19). *Gender agreement and semantic processing in French children: ERP effects of age and proficiency*. https://doi.org/10.13140/RG.2.2.28154.59840

Courteau, É., Loignon, G., Royle, P., & Steinhauer, K. (Submitted). Identifying Linguistic Markers of French-speaking Teenagers With Developmental Language Disorder: Which Tasks Matter? *Journal of Speech, Language and Hearing Research*.

Courteau, É., Martignetti, L., Royle, P., & Steinhauer, K. (2019). Eliciting ERP components for morphosyntactic agreement mismatches in perfectly grammatical sentences. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.01152

Courteau, É., Royle, P., Gascon, A., Marquis, A., Drury, J. E., & Steinhauer, K. (2013). Gender concord and semantic processing in french children: An auditory ERP study. In S. Baiz, N. Goldman, & R. Hawkes (Eds.), *Proceedings of the 37th annual BUCLD* (Vol. 1, pp. 87–99).

Courteau, É., Steinhauer, K., & Royle, P. (2015). L'acquisition du groupe nominal en français et de ses aspects morpho-syntaxiques et sémantiques: Une étude depotentiels évoqués. *Glossa*, *117*, 77–93.

Cummings, A., Čeponienė, R., Dick, F., Saygin, A. P., & Townsend, J. (2008). A developmental ERP study of verbal and non-verbal semantic processing. *Brain Research*, *1208*, 137–149. https://doi.org/10.1016/j.brainres.2008.02.015

Daniel, T. A., Katz, J. S., & Robinson, J. L. (2016). Delayed match-to-sample in working memory: A BrainMap meta-analysis. *Biological Psychology*, *120*, 10–20. https://doi.org/10.1016/j.biopsycho.2016.07.015

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Dube, S., Kung, C., Brock, J., & Demuth, K. (2018). Perceptual salience and the processing of subject-verb agreement in 9-11-year-old English-speaking children: Evidence from ERPs. *Language Acquisition*. https://doi.org/10.1080/10489223.2017.1394305

Dube, S., Kung, C., Peter, V., Brock, J., & Demuth, K. (2016). Effects of type of agreement violation and utterance position on the auditory processing of subject-verb agreement: An ERP study. *Frontiers in Psychology*, *7*, 1276. https://doi.org/10.3389/fpsyg.2016.01276

Durand, J., & Lyche, C. (2008). French liaison in the light of corpus data. *Journal of French Language Studies*, *18*(1), 33–66. https://doi.org/10.1017/S0959269507003158

Franck, J., Cronel-Ohayon, S., Chillier, L., Frauenfelder, U. H., Hamann, C., Rizzi, L., & Zesiger, P. (2004). Normal and pathological development of subject–verb agreement in speech production: A study on French children. *Journal of Neurolinguistics*, *17*(2), 147–180. https://doi.org/10.1016/S0911-6044(03)00057-5

Friedrich, M., & Friederici, A. D. (2004). N400-like semantic incongruity effect in 19-month-olds: Processing known words in picture contexts. *Journal of Cognitive Neuroscience*, *16*(8), 1465–1477. MEDLINE. https://doi.org/10.1162/0898929042304705

Friedrich, M., & Friederici, A. D. (2006). Early N400 development and later language acquisition. *Psychophysiology*, *43*(1), 1–12. MEDLINE. https://doi.org/10.1111/j.1469-8986.2006.00381.x

Fromont, L. A., Steinhauer, K., & Royle, P. (2020). Verbing nouns and nouning verbs: Using a balanced design provides ERP evidence against “syntax-first” approaches to sentence processing. *PLOS ONE*, *15*(3), e0229169. https://doi.org/10.1371/journal.pone.0229169

Haebig, E., Weber, C., Leonard, L. B., Deevy, P., & Tomblin, J. B. (2017). Neural patterns elicited by sentence processing uniquely characterize typical development, SLI recovery, and SLI persistence. *Journal of Neurodevelopmental Disorders*, *9*(1), 22. https://doi.org/10.1186/s11689-017-9201-1

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

Herbay, A., & Steinhauer, K. (2020, October 5). *ERPscope: A new R package to easily visualize ERP data*. Live MEEG, Online. https://livemeeg2020.org/wp-content/uploads/2020/10/LiveMEEG2020_posters.pdf

Huang, Y., & Ferreira, F. (2020). The Application of Signal Detection Theory to Acceptability Judgments. *Frontiers in Psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.00073

Isel, F., & Kail, M. (2018). Morphosyntactic integration in French sentence processing: Event-related brain potentials evidence. *Journal of Neurolinguistics*, *46*, 23–36. https://doi.org/10.1016/j.jneuroling.2017.12.006

Jakubowicz, C. (2003). Hypothèses psycholinguistiques sur la nature du déficit dysphasique. In C.-L. Gérard & V. Brun (Eds.), *Les dysphasies.* (pp. 23–70). Masson.

Juottonen, K., Revonsuo, A., & Lang, H. (1996). Dissimilar age influences on two ERP waveforms (LPC and N400) reflecting semantic context effect. *Brain Research. Cognitive Brain Research*, *4*(2), 99–107. https://doi.org/10.1016/0926-6410(96)00022-5

Kail, R. (1994). A Method for Studying the Generalized Slowing Hypothesis in Children With Specific Language Impairment. *Journal of Speech, Language, and Hearing Research*, *37*(2), 418–421. https://doi.org/10.1044/jshr.3702.418

Kail, R., & Salthouse, T. A. (1994). Processing speed as a mental capacity. *Acta Psychologica*, *86*(2), 199–225. https://doi.org/10.1016/0001-6918(94)90003-5

Kandylaki, K. D., & Bornkessel-Schlesewsky, I. (2019). From story comprehension to the neurobiology of language. *Language, Cognition and Neuroscience*, *34*(4), 405–410. https://doi.org/10.1080/23273798.2019.1584679

Katsos, N., Roqueta, C. A., Estevan, R. A. C., & Cummins, C. (2011). Are children with Specific Language Impairment competent with the pragmatics and logic of quantification? *Cognition*, *119*(1), 43–57. https://doi.org/10.1016/j.cognition.2010.12.004

Kornilov, S. A., Magnuson, J. S., Rakhlin, N., Landi, N., & Grigorenko, E. L. (2015). Lexical processing deficits in children with developmental language disorder: An event-related potentials study. *Development and Psychopathology*, *27*(2), 459–476. https://doi.org/10.1017/S0954579415000097

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*, 621–647. 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

Lachaud, C. M., & Renaud, O. (2011). A tutorial for analyzing human reaction times: How to filter data, manage missing values, and choose a statistical model. *Applied Psycholinguistics*, *32*(2), 389–416.

Leclercq, A.-L., Quémart, P., Magis, D., & Maillart, C. (2014). The sentence repetition task: A powerful diagnostic tool for French children with specific language impairment. *Research in Developmental Disabilities*, *35*(12), 3423–3430. https://doi.org/10.1016/j.ridd.2014.08.026

Legendre, G., Culbertson, J., Zaroukian, E., Hsin, L., Barrière, I., & Nazzi, T. (2014). Is children's comprehension of subject–verb agreement universally late? Comparative evidence from French, English, and Spanish. *Lingua*, *144*, 21–39.

Leonard, L. B., Fey, M. E., Deevy, P., & Bredin-Oja, S. L. (2015). Input sources of third person singular -s inconsistency in children with and without specific language impairment*. *Journal of Child Language*, *42*(4), 786–820. https://doi.org/10.1017/S0305000914000397

Leonard, L. B., Weismer, S. E., Miller, C. A., Francis, D. J., Tomblin, J. B., & Kail, R. (2007). Speed of processing, working memory, and language impairment in children. *Journal of Speech, Language, and Hearing Research*, *50*(2), 408–428. https://doi.org/10.1044/1092-4388(2007/029)

Lété, B., Sprenger-Charolles, L., & Colé, P. (2004). MANULEX: A grade-level lexical database from French elementary school readers. *Behavior Research Methods, Instruments, & Computers*, *36*(1), 156–166. https://doi.org/10.3758/BF03195560

Luck, S. J., Kappenman, E. S., Fuller, R. L., Robinson, B., Summerfelt, A., & Gold, J. M. (2009). Impaired response selection in schizophrenia: Evidence from the P3 wave and the lateralized readiness potential. *Psychophysiology*, *46*(4), 776–786. https://doi.org/10.1111/j.1469-8986.2009.00817.x

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide, 2nd ed*. Lawrence Erlbaum Associates Publishers.

Marquis, A., & Royle, P. (2019). Verb acquisition in monolingual and multilingual children and adults. In P. Guijarro-Fuentes & C. Suárez-Gómez (Eds.), *Proceedings of the GALA: Language Acquisition and Development* (Newcastle upon Tyne, pp. 307–324).

Cambridge Scholars Publishing. https://www.researchgate.net/publication/332471519_Verb_acquisition_in_monolingual_and_multilingual_children_and_adults/link/5cb76acca6fdcc1d499c455b/download

Miller Carol A., Leonard Laurence B., Kail Robert V., Zhang Xuyang, Tomblin J. Bruce, & Francis David J. (2006). Response Time in 14-Year-Olds With Language Impairment. *Journal of Speech, Language, and Hearing Research*, *49*(4), 712–728. https://doi.org/10.1044/1092-4388(2006/052)

Misirliyan, C., Courteau, E., & Royle, P. (2020, January 31). *Correction des clignements oculaires dans les électroencéphalogrammes (EEG) de jeunes ayant un trouble du langage [Correcting eye-blink artefacts in EEG from children with language disorders]*. 53e Congrès Premier des stagiaires de recherche du 1er cycle de la Faculté de médecine, Montréal, QC. https://doi.org/10.13140/RG.2.2.14050.25289

Molinaro, N., Barber, H. A., & Carreiras, M. (2011). Grammatical agreement processing in reading: ERP findings and future directions. *Cortex*, *47*(8), 908–930. https://doi.org/10.1016/j.cortex.2011.02.019

Montgomery, J. W., Gillam, R. B., & Evans, J. L. (2016). Syntactic versus memory accounts of the sentence comprehension deficits of specific language impairment: Looking back, looking ahead. Journal of Speech, Language, and Hearing Research, 59(6), 1491-1504.

Morgan, E. U., van der Meer, A., Vulchanova, M., Blasi, D. E., & Baggio, G. (2020). Meaning before grammar: A review of ERP experiments on the neurodevelopmental origins of semantic processing. *Psychonomic Bulletin & Review*, *27*(3), 441–464. https://doi.org/10.3758/s13423-019-01677-8

Neville, H. J., Coffey, S. A., Holcomb, P. J., & Tallal, P. (1993). The Neurobiology of Sensory and Language Processing in Language-Impaired Children. *Journal of Cognitive Neuroscience*, *5*(2), 235–253. https://doi.org/10.1162/jocn.1993.5.2.235

New, B., Pallier, C., Ferrand, L., & Matos, R. (2001). Une base de données lexicales du français contemporain sur internet: LEXIQUE™//A lexical database for contemporary french: LEXIQUE™. *L'année Psychologique*, *101*(3), 447–462. https://doi.org/10.3406/psy.2001.1341

Noveck, I. A. (2001). When children are more logical than adults: Experimental investigations of scalar implicature. *Cognition*, *78*(2), 165–188. https://doi.org/10.1016/S0010-0277(00)00114-1

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Computational Intelligence and Neuroscience*, *2011*, 9. https://doi.org/doi.org/10.1155/2011/156869

Osterhout, L., & Mobley, L. A. (1995). Event-related brain potentials elicited by failure to agree. *Journal of Memory and Language*, *34*(6), 739–773. https://doi.org/10.1006/jmla.1995.1033

Papafragou, A., & Musolino, J. (2003). Scalar implicatures: Experiments at the semantics–pragmatics interface. *Cognition*, *86*(3), 253–282. https://doi.org/10.1016/S0010-0277(02)00179-8

Pawlowska, M., Robinson, S., & Seddoh, A. (2014). Detection of Lexical and Morphological Anomalies by Children With and Without Language Impairment. *Journal of Speech,* *Language, and Hearing Research*, *57*(1), 236–246. https://doi.org/10.1044/1092-4388(2013/12-0241)

Pijnacker, J., Davids, N., van Weerdenburg, M., Verhoeven, L., Knoors, H., & van Alphen, P. (2017). Semantic processing of sentences in preschoolers with specific language impairment: Evidence from the N400 effect. *Journal of Speech, Language, and Hearing Research*, *60*(3), 627–639. https://doi.org/10.1044/2016_JSLHR-L-15-0299

Pourquié, M. (2015). fLEX: Multilingual assessment of inFlectional and LEXical processing. *Software, Intellectual Property 2016-01*, *88*.

Pourquié, M., Courteau, É., Duquette, A.-S., & Royle, P. (in revision). *Verb inflection and argument structure processing in French adolescents with DLD.*

Purdy, J. D., Leonard, L. B., Weber-Fox, C., & Kaganovich, N. (2014). Decreased sensitivity to long-distance dependencies in children with a history of specific language impairment: Electrophysiological evidence. *Journal of Speech, Language, and Hearing Research*, *57*(3), 1040–1059. https://doi.org/10.1044/2014_JSLHR-L-13-0176

Quémart, P., & Maillart, C. (2016). The sensitivity of children with SLI to phonotactic probabilities during lexical access. *Journal of Communication Disorders*, *61*, 48–59. https://doi.org/10.1016/j.jcomdis.2016.03.005

Redmond Sean M., Ash Andrea C., & Hogan Tiffany P. (2015). Consequences of Co-Occurring Attention-Deficit/Hyperactivity Disorder on Children's Language Impairments. *Language, Speech, and Hearing Services in Schools*, *46*(2), 68–80. https://doi.org/10.1044/2014_LSHSS-14-0045

Regel, S., Opitz, A., Müller, G., & Friederici, A. D. (2015). The Past Tense Debate Revisited: Electrophysiological Evidence for Subregularities of Irregular Verb Inflection. *Journal of Cognitive Neuroscience*, *27*(9), 1870–1885. https://doi.org/10.1162/jocn_a_00818

Rietveld, T., & van Hout, R. (2015). The t test and beyond: Recommendations for testing the central tendencies of two independent samples in research on speech, language and hearing pathology. *Journal of Communication Disorders*, *58*, 158–168. https://doi.org/10.1016/j.jcomdis.2015.08.002

Royle, P., & Courteau, É. (2014). Language processing in children with specific language impairment: A review of event-related potential studies. In L. T. Klein & V. Amato (Eds.), *Language processing: New research* (pp. 33–64). Nova Science Publishers.

Royle, P., Drury, J. E., & Steinhauer, K. (2013). ERPs and task effects in the auditory processing of gender agreement and semantics in French. *The Mental Lexicon*, *8*(2), 216–244. https://doi.org/10.1075/ml.8.2.05roy

Royle, P., Jarema, G., & Kehayia, E. (2002). Auditory Verb Recognition in Developmental Language Impairment. *Brain and Language*, *81*(1), 487–500. https://doi.org/10.1006/brln.2001.2541

Royle, P., & Steinhauer, K. (in preparation). *Neural correlates of morphology computation and representation* (p. in press).

Schneider, J. M., & Maguire, M. J. (2019). Developmental differences in the neural correlates supporting semantics and syntax during sentence processing. *Developmental Science*, *22*(4), e12782. https://doi.org/10.1111/desc.12782

Secord, W. A., Wiig, E., Boulianne, L., Semel, E., & Labelle, M. (2009). *Évaluation clinique des notions langagières fondamentales®—Version pour francophones du Canada (CELF® CDN-F)*. The Psychological Corporation.

Steinhauer, K., & Connolly, J. F. (2008). Event-related potentials in the study of language. In B. Stemmer (Ed.), *Handbook of the neuroscience of language* (pp. 91–104). Elsevier Ltd.

Steinhauer, K., & Ullman, M. T. (2002). Consecutive ERP effects of morpho-phonology and morpho-syntax. *Brain and Language*, *83*, 62–65.

Steinhauer, K., White, E. J., & Drury, J. E. (2009). Temporal dynamics of late second language acquisition: Evidence from event-related brain potentials. *Second Language Research*, *25*(1), 13–41. https://doi.org/10.1177/0267658308098995

Stowe, L. A., Kaan, E., Sabourin, L., & Taylor, R. C. (2018). The sentence wrap-up dogma. *Cognition*, *176*, 232–247. https://doi.org/10.1016/j.cognition.2018.03.011

Tippmann, J., Stärk, K., Ebersberg, M., Opitz, A., & Rossi, S. (2018). Developmental changes in neuronal processing of irregular morphosyntactic rules during childhood. *The Nijmegen Lectures 2018*.

Ullman, M. T. (2020). The Declarative/Procedural Model: A Neurobiologically-Motivated Theory of First and Second Language. In B. VanPatten, G. D. Keating, & S. Wulff (Eds.), *Theories in Second Language Acquisition: An Introduction* (3rd ed., pp. 128–161). Routledge.

Ullman, M. T. (2001). The neural basis of lexicon and grammar in first and second language: The declarative/procedural model. Bilingualism: Language and cognition, 4(2), 105-122.

Ullman, M. T., Earle, F. S., Walenski, M., & Janacsek, K. (2020). The Neurocognition of Developmental Disorders of Language. *Annual Review of Psychology*, *71*(1), null. https://doi.org/10.1146/annurev-psych-122216-011555

Ullman, M. T., & Pierpont, E. I. (2005). Specific language impairment is not specific to language: The procedural deficit hypothesis. *Cortex*, *41*(3), 399–433. https://doi.org/10.1016/S0010-9452(08)70276-4

Valois, D., Royle, P., Sutton, A., & Bourdua-Roy, È. (2009). L'ellipse du nom en français: Le rôle des données de l'acquisition pour la théorie linguistique. *The Canadian Journal of Linguistics/La Revue Canadienne de Linguistique*, *54*(2), 339–366.

van der Lely, H. K. J. (1998). SLI in Children: Movement, Economy, and Deficits in the Computational-Syntactic System. *Language Acquisition*, *7*(2–4), 161–192. https://doi.org/10.1207/s15327817la0702-4_4

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

Weber-Fox, C., Leonard, L. B., Wray, A. H., & Tomblin, J. B. (2010). Electrophysiological correlates of rapid auditory and linguistic processing in adolescents with specific language impairment. *Brain and Language*, *115*(3), 162–181. https://doi.org/10.1016/j.bandl.2010.09.001

Windsor, J., & Hwang, M. (1999). Testing the Generalized Slowing Hypothesis in Specific Language Impairment. *Journal of Speech, Language, and Hearing Research*, *42*(5), 1205–1218. https://doi.org/10.1044/jslhr.4205.1205

Wobbrock, J. O., Findlater, L., Gergle, D., & Higgins, J. J. (2011). The aligned rank transform for nonparametric factorial analyses using only anova procedures. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 143–146. https://doi.org/10.1145/1978942.1978963

Zhang, J., & Mueller, S. T. (2005). A note on ROC analysis and non-parametric estimate of sensitivity. *Psychometrika*, *70*(1), 1–10. https://doi.org/10.1007/s11336-003-1119-8

Ziegler, J. C., Castel, C., Pech-Georgel, C., George, F., Alario, F.-X., & Perry, C. (2008). Developmental dyslexia and the dual route model of reading: Simulating individual differences and subtypes. *Cognition*, *107*(1), 151–178. https://doi.org/10.1016/j.cognition.2007.09.004

6.6. Supplementary material manuscript 3

6.6.1. Section 1. Behavioural performance

Supplementary Table 1

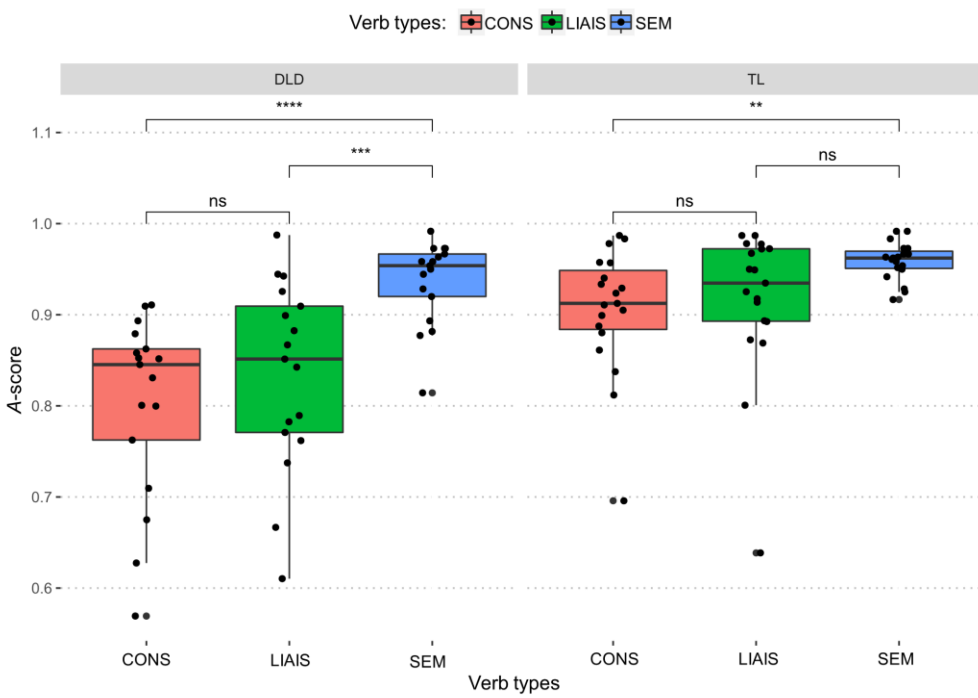
Acceptability judgements for match and mismatch conditions.

Conditions	DLD group Match	DLD group Mismatch	TL group Match	TL group Mismatch
Semantics	84.7 (0.12) *30*	93.7 (.063) *30*	88.8 (0.07) *30*	96.0 (0.04) *30*
SUBJECT CONTEXT	84.7 (0.21) *15*	92.3 (0.09) *15*	93.7 (0.07) *15*	95.3 (0.06) *15*
NEUTRAL CONTEXT	84.7 (0.12) *15*	95.3 (0.08) *15*	84.0 (0.12) *15*	96.7 (0.04) *15*
Consonant-final verbs	85.1 (0.12) *60*	60.3 (0.18) *60*	91.2 (.06) *60*	77.8 (0.15) *60*
SG: SUBJECT CONTEXT	83.1 (0.24) *15*	67.5 (0.21) *15*	92.0 (.01) *15*	77.3 (0.17) *15*
SG: NEUTRAL CONTEXT	85.1 (0.09) *15*	39.2 (0.22) *15*	91.6 (.09) *15*	67.0 (0.29) *15*
PL: SUBJECT CONTEXT	86.0 (0.23) *15*	78.0 (0.20) *15*	92.7 (.06) *15*	85.3 (0.01) *15*
PL: NEUTRAL CONTEXT	86.3 (0.10) *15*	57.3 (0.25) *15*	88.3 (.10) *15*	81.7 (0.13) *15*
Liaison verbs	83.23 (0.23) *60*	63.4 (0.19) *60*	93.4 (.05) *60*	78.4 (0.19) *60*
SG: SUBJECT CONTEXT	88.8 (0.12) *15*	69.7 (0.26) *15*	93.7 (.07) *15*	85.3 (0.17) *15*
SG: NEUTRAL CONTEXT	91.0 (0.10) *15*	44.0 (0.28) *15*	93.0 (.10) *15*	71.7 (0.22) *15*
PL: SUBJECT CONTEXT	88.6 (0.21) *15*	80.0 (0.21) *15*	95.6 (.06) *15*	81.2 (0.19) *15*
PL: NEUTRAL CONTEXT	88.6 (0.08) *15*	57.3 (0.28) *15*	91.3 (.10) *15*	75.3 (0.23) *15*

Note. Accuracy means (and standard deviations) for audio-visually matching and mismatching trials in lexico-semantic and number conditions for both consonant-final and liaison morphosyntactic conditions.

Supplementary Figure 1

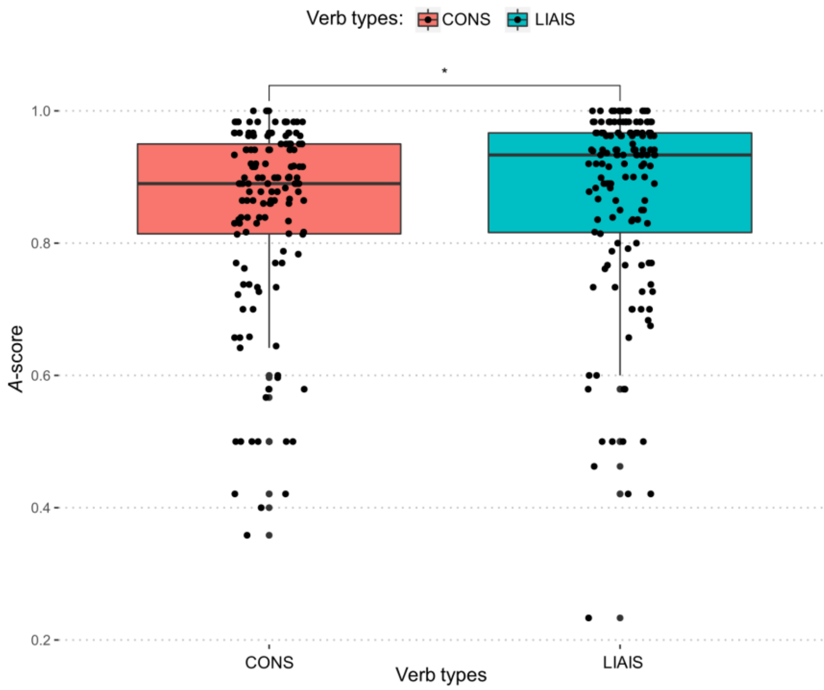
Boxplots illustrating the G*ROUP* × V*ERB interaction*



Note. CONS = Consonant-final verbs; LIAIS = Liaison verbs; SEM = Lexico-semantic verbs; DLD = DLD group; TL = Typical language group. *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$, and ****: $p < 0.0001$

Supplementary Figure 2

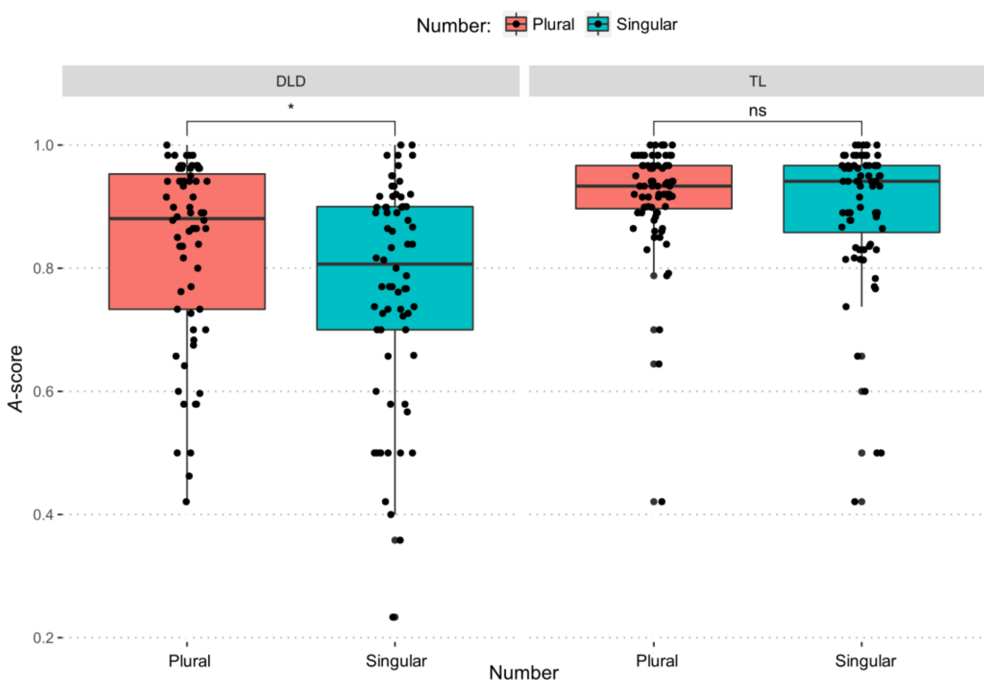
Boxplots illustrating the main effect of V*ERB for morphosyntactic conditions.*



Note. CONS = Consonant-final verbs; LIAIS = Liaison verbs. *: $p < 0.05$

Supplementary Figure 3

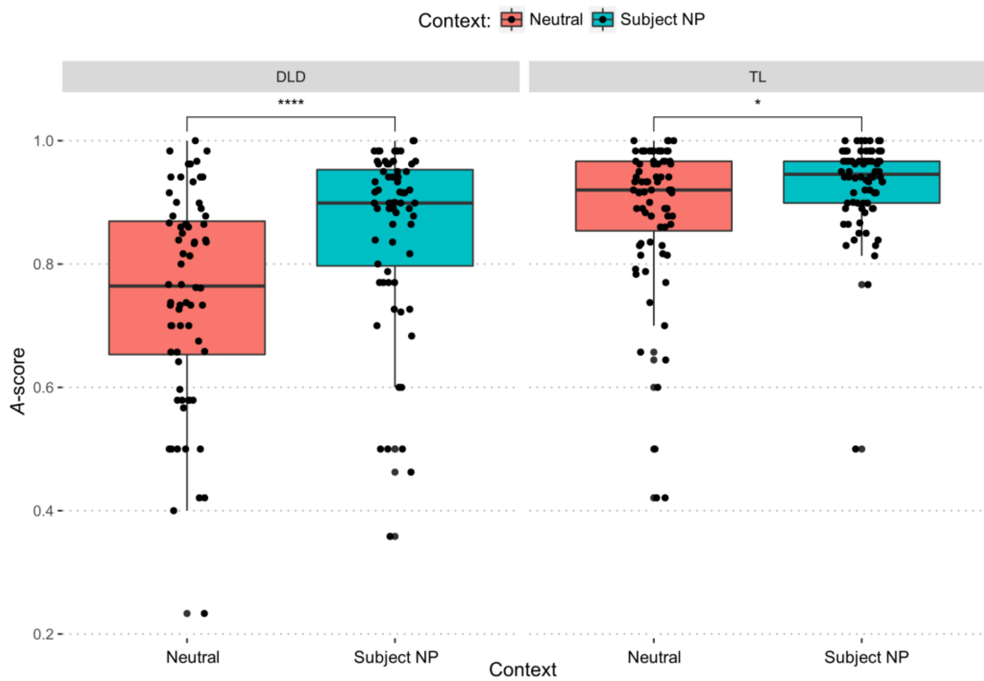
Boxplots illustrating the GROUP × NUMBER interaction.



Note. DLD = DLD group; TL = Typical language group; ns = nonsignificant. *: $p < 0.05$

Supplementary Figure 4

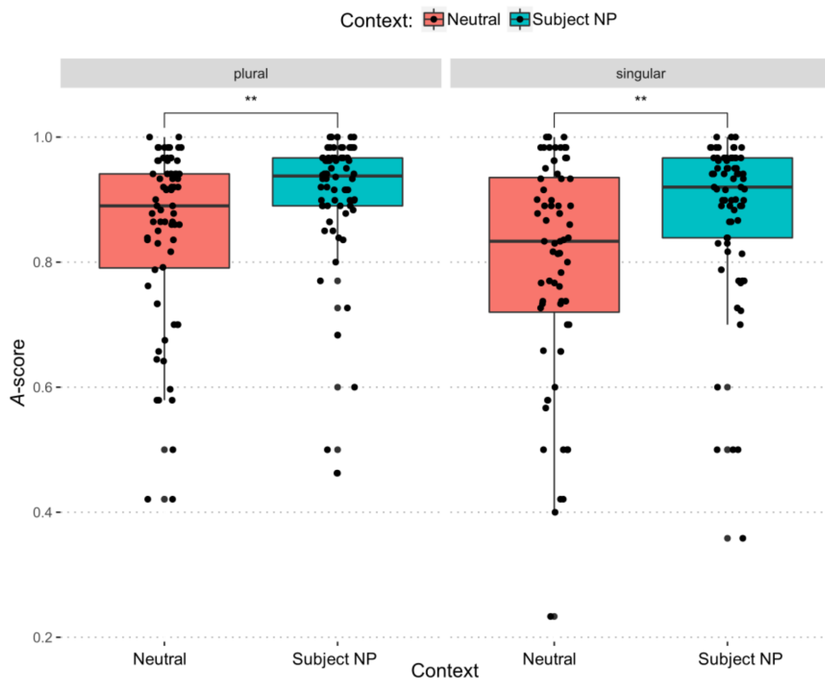
Boxplots illustrating the GROUP × CONTEXT interaction.



Note. DLD = DLD group; TL = Typical language group. *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$, and ****: $p < 0.0001$

Supplementary Figure 5

Boxplots illustrating the CONTEXT × NUMBER interaction.



Note. **: $p < 0.01$

6.6.2. Section 2. Problematic effects for the sentence onset conditions

Recall that participants heard two kinds of contexts at sentence onset: neutral contexts (naming a general characteristic of the depicted scene e.g., 'each week') or subject NP contexts, where the NP either matched or not with the picture in terms of number agreement at the determiner (le/la/les 'the. M.$_{\text{SG/F. SG/PL}}$'). At this point in the sentence, match and mismatch neutral contexts conditions were identical, the mismatches occurring only at verb onset further downstream.

We present here the problematic effects for the sentence onset condition: refer to Table 3 of the main article for the complete statistical analyses. In the 400–600 TW, the decomposition of effects showed significant results for the DLD group only. In the lateral electrodes, follow-up analyses on the GROUP × CONTEXT × CONDITION interaction revealed a significant effect of CONDITION for subject contexts in the DLD group. In midline channels we also found a significant effect of CONDITION which was constrained to the Fz electrode, but this time in neutral contexts for the DLD group, as revealed by the significant GROUP × CONTEXT × ELECTRODE × CONDITION interaction decomposition. In the 1000–1200 ms time-window, the DLD group showed a significant effect of CONDITION for the neutral contexts, as indicated by the GROUP × HEMISPHERE × CONTEXT × CONDITION and GROUP × CONTEXT × ELECTRODE × CONDITION interactions in respectively lateral and midline channels. The fact that the neutral context effect was not driven by our stimuli and that none of these effects were found in the TL or the adults in the same experiment prompted us to check if this pattern was shared by a majority of DLD participants. Visual inspection of the neutral condition revealed that three participants (D09, D11 and D12) showed large positivities between the mismatch and match conditions in neutral contexts at sentence onset. To see if the peculiar effects were still present without these participants, we conducted global ANOVAs with the reduced DLD group (n = 13). Following our expectations, these effects did not reach significance in the interaction decomposition (see right columns for DLD = 13 in Table 2), showing that these effects were mainly driven by these 3 participants. See Table 2 for a comparison of the problematic effects between the complete DLD group (N = 16) and the subset group without the participants with divergent patterns (n = 13).

Supplementary Table 2

Global ANOVAs for Sentence onset's problematic effects at verb onset at time-windows of interest.

	DLD N = 16			DLD n = 13		
	df	400–600	1000–1200	*df*	400–600	1000–1200
LATERAL ELECTRODES						
GROUP × CONT × COND	(1,33)	7.28**	—	(1,30)	5.02*	—
DLD: CONT × COND	(1,15)	7.13*	—	(1,12)	3.77†	—
DLD: NP: COND	(1,15)	4.44*	—		—	—
GROUP × HEMI × CONT × COND	(1,33)	—	4.78*	(1,30)	—	6.76*
RIGHT HEMI : GROUP × CONT × COND	(1,33)	—	5.42*	(1,30)	—	5.13*
RIGHT HEMI : NEUT : GROUP × COND	(1,33)	—	6.58*	(1,30)	—	4.70*
RIGHT HEMI : NEUT : DLD: COND	(1,15)	—	6.84*		—	3.74 †
MIDLINE ELECTRODES						
GROUP × CONT × COND	(1,33)	6.82**	—	(1,30)	4.66*	—
DLD: CONT × COND	(1,15)	9.07**	—	(1,12)	4.93*	—
DLD: NP: COND	(1,15)	3.96†	—	(1,12)	3.92†	—
DLD: NEUT: COND	(1,15)	6.57*	—		—	—
GROUP × CONT × ELECTRODE × COND	(3,99)	—	4.57*	(3,90)	—	3.95*
FZ: CONT × GROUP × COND	(1,33)	—	8.99**	(1,30)	—	6.46*
FZ: NEUT : GROUP × COND	(1,33)	—	5.38*	(1,30)	—	3.17 †
FZ: NEUT : DLD : COND	(1,15)	—	4.59*		—	—

Note. Only significant results and trends are presented. Cond = Condition; Cont = Context; Neut = neutral context; NP = Subject NP context; Numb = Number; SG = Singular; Ant = Anteriority; Lat = Laterality; Hemi = Hemisphere. †: $p < 0.10$, *: $p < 0.05$, **: $p < 0.01$, and ***: $p < 0.001$.

6.6.3. Section 3. Response contingency analyses and ERP amplitude for consonant-final plural conditions

The CONS plural verbs condition elicited a N400-like negativity in the 500–800 ms time-window in both groups. This result raised the question as whether the participant's ability to discriminate mismatch from match was directly reflected by ERP measures. We investigated this relationship in two ways, in trial-based and participant-based analyses.

In trial-based analyses, single-subject ERP averages can be "response contingent" that is based on correct behavioral trials or on "all trials" irrespective of accuracy (White et al., 2012).

By comparing the grand average ERPs of all trials and correct-only trials, we checked whether participants engaged the same neurocognitive processes for correctly versus incorrectly judged sentences. To make this comparison, first we ran a global ANOVA for the midline channels of correct-only trials ERPs (see Table 3 below). We found a significant ELECTRODE × CONDITION interaction and its decomposition revealed a significant effect of CONDITION at PZ and OZ. Theses results confirmed the N400-like negativity for the correct-only trials. However, the distribution was in centro-occipital regions when compared to the effect in all (correct and incorrect) trials, where we found a main effect of CONDITION and no interaction with the topographic factor ELECTRODE. Following these analyses, we wondered if the amplitude of the effect of all trials differed statically from the one in correct-trials only. To do so, we choose the Pz channels since it was where we found the strongest effect in correct-trials-only ANOVAs. We compared the effect at Pz directly by conducting a repeated-measures ANOVA on the mean amplitude of the N400-like difference wave (i.e., mean amplitude of mismatch sentences minus mean amplitude of match sentences) using ANALYSIS TYPE (two levels: all trials vs. correct only) as the within-subject factor and GROUP (DLD, TL) as a between-subject factor. We found a trend for the main effect of ANALYSIS TYPE ($F(1,33) = 2.83$, $p < 0.10$), and no interaction.

Supplementary Table 3

Global ANOVAs for the CONS plural verbs response contingency analyses.

		N400-like negativity 500–800 ms	
	df	All trials	Correct-Only Trials
MIDLINE ELECTRODES			
CONDITION	(1,33)	10.08**	3.75†
ELECTRODE × COND	(3,99)	—	3.75*
CZ: COND	(1,34)	—	3.76†
PZ: COND	(1,34)	—	8.05**
OZ: COND	(1,34)	—	4.91*

Note. Only significant results and trends are presented. Cond = Condition. †: $p < 0.10$, *: $p < 0.05$, **: $p < 0.01$, and ***: $p < 0.001$

We observed that both all trials and correct-only trials analyses revealed a N400-like negativity, with similar amplitude. However, we know that children and teenagers demonstrate considerable variability in their behaviour, and this could be reflected in their ability to process mismatches and identify them. To investigate the relationship between individual learner's behavioural performance and N400-like effects, we ran bivariate correlation between participants' judgment accuracy (*A*-scores) and the mean amplitude of the N400-like negativity difference wave for all trials at a representative channel (Pz). We found no significant relationship between individual ERP results and behavioural performance (excluding 3 outliers participants with component amplitude > 10μV : $r(30) = 0.22$, $p = 0.25$). Our results pointed toward an inverse relationship: the lower the μV, the lower the *A*-score. We also ran the same analyses including two other variables: first, age and mean amplitude at Pz, and second, score on sentence repetition and mean amplitude at Pz. No significant correlations were found.

To summarize, the participants' ability to identify mismatches did not emerge as being related to ERPs in the CONS plural verb condition. In trial-based analyses, response contingency analyses revealed an N400-like negativity for correct-trials only, which had a more centro-occipital distribution than the all-trials effect, but which did not differ in terms of amplitude. In participant-based analyses, we did not find significant correlations between their ability to detect mismatches and ERP amplitude, nor between their age or their score on sentence repetition and the ERP amplitude. Note that the CONS plural condition was the most logical condition to test this effect because it was the morphosyntactic condition that elicited a similar ERP pattern in both groups.

Considering that we were not able to demonstrate a relationship between participants' ability to discriminate mismatch from match CONS plural sentences and ERP measures, we did not pursue these analyses in the subsequent analyses.

6.6.4. Section 4. Analyses of the positivity in Consonant-Final singular conditions

Supplementary Table 4

Global ANOVAs for Consonant-Final singular verbs comparing directly the 1000–1400 and 1400–1800 ms time-windows.

	df	
LATERAL ELECTRODES		
CONDITION	(1,33)	5.64*
TIME WINDOW	(1,33)	4.81*
TW × GROUP × COND		3.88 †
TW × HEMI × COND	(1,33)	6.83*
TW × ANT × HEMI × COND	(2,66)	6.75**
1400-1800: ANT × HEMI × COND	(1,34)	4.96*
1400-1800: FRONTAL : HEMI X COND	(1,34)	4.85*
1400-1800: FRONTAL: RIGHT HEMI: COND	(1,34)	11.2**
TW × LAT × HEMI × COND	(1,33)	5.60*
LATERAL: TW × HEM × COND	(1,34)	6.90**
LATERAL: RIGHT HEMI: TW × COND	(1,34)	3.32 †
TW × ANT × LAT × HEM × COND	(2,66)	6.60**
FRONTAL: TW × LAT × HEM × COND	(1,34)	7.78**
FRONTAL: LEFT HEMI : TW × LAT × COND	(1,34)	10.29**
FRONTAL: LEFT HEMI : LAT : TW × COND	(1,34)	3.25†
TW × GROUP × LAT × COND	(1,33)	6.97*
TL: TW × LAT × COND	(1,18)	4.11†
DLD: TW × LAT × COND	(1,15)	3.66†
GROUP × HEMI × LAT × COND	(1,33)	4.58*
MIDLINE ELECTRODES		
CONDITION	(1,4)	5.85*
TIME WINDOW	(1,33)	—
TIME WINDOW × GROUP × COND		5.33*
1000-1800: GROUP × COND	(1,33)	—
1000-1400: GROUP × COND	(1,33)	7.42**
1000-1400: DLD: COND	(1,15)	—
1000-1400: TL: COND	(1,18)	7.30*

Note. Only significant results and trends are presented. Cond = Condition; Ant = Anteriority; Lat = Laterality; Hemi = Hemisphere. †: $p < 0.10$, *: $p < 0.05$, **: $p < 0.01$, and ***: $p < 0.001$.

6.6.5. Section 5. Refinement of the PDH predictions for DLD ERP patterns

In this section, we present more detailed predictions of the PDH for the DLD ERP patterns. First, we present general predictions from the PDH from the perspective of the preserved declarative and impaired procedural memory systems. Then, we present specific predictions for DLD ERP patterns for each experimental condition in relation to the results obtained with adults in the same experiment (Courteau et al., 2019).

Predictions for the declarative memory:

- Since the declarative-procedural model (DP; Ullman, 2004, 2020) predicts that lexical memory depends on the declarative memory and that the PDH predicts that this system should remain largely normal in DLD, we expect that lexical manipulations should consistently elicit N400s in TL and DLD participants.
- The iterature suggests that lexico-semantic processing as reflected by the N400 can be seen in children starting at 2 years old, and that N400s similar to adults in terms of latency and distribution can be elicited in children starting at 7 years old (see section 6.1.2). Considering this and that the declarative memory system improves in childhood and is well developed in adolescence (Ullman, 2020), we thus expect that both the DLD and TL groups, aged from 8 to 15 years old, will elicit N400s reflecting declarative memory processing.

Predictions for the procedural memory:

- Grammatical disruption of morphosyntax, such as agreement, often produces two components. First ANs which are expected to reflect, at least partly, procedural memory-based processes (Ullman, 2001). This component is often followed by a P600, which is not thought to reflect procedural processing (Ullman, 2020). Note that is some studies, agreement has been known to elicit P600s only (e.g., Osterhout and Mobley, 1995).

- It is expected that participants from the TL group have a normally developing procedural memory system, and this system should already be well developed in early childhood and will be used to learn grammar (Ullman, 2020). Based on this, the TL group should exhibit ANs in response to morphosyntactic disruptions. If processing is not mature in TL and they are still acquiring the linguistic materials included in an experimental condition (e.g., consonant-final SUBIRR verbs, see results of manuscript #1) , we predict that the ERPs associated with automatised procedural memory processing, i.e, (sustained) AN, would not be elicited.
- Since the procedural system is expected to be impaired in DLD, grammatical processing could be supported by the declarative one as a compensation mechanism, and grammatical disruption may elicit N400s in DLD instead of ANs.

Supplementary Table 5

Predictions for each condition based on adults' ERP in the same experimental conditions

Linguistic materials and experimental conditions	Expected ERPs for TL based on adults ERP patterns (Courteau et al., 2019)	Declarative or procedural processing based on the adults' ERPs and the DP model	Expected ERPs for the DLD group based on the PDH and the adults' results
Mismatches on verbs reflecting lexico-semantic processing	N400 (with additional frontal and sentence-final negativities)	Declarative memory system processing. Since the condition targets lexical comprehension and not retrieval, and that participants did not have a limited time to give their acceptability judgement, this task is not expected to tap into procedural processing (Ullman and Pierpont, 2005).	N400: declarative memory system processing
Agreement on determiner in NP at sentence onset reflecting regular morphosyntactic agreement	N400-P600	The N400 suggests declarative memory system processing. The P600 is not associated directly with procedural processing (Ullman, 2020). However, the DP model suggests that when agreement is made between constituents, it taps into rule-governed morphosyntax (Steinhauer and Ullman, 2002),and thus the procedural memory system should play a role. However, this is not illustrated clearly in the adults ERP patterns of this condition, because no AN was elicited.	The DP model suggests that when agreement is done between constituents, the procedural memory system should play a role. Considering this, and the PDH, we would expect in DLD at least partially different ERP patterns indicating the compensation of the impaired procedural system by the declarative one.
Liaison verbs used to test	Early sustained (left) AN, and a P600	This sustained (left) AN suggests automatized morphosyntactic	The PDH predicts deficits of procedural memory processing

regular morphosyntactic agreement		processing which is subserved by the procedural memory system. This is coherent with the DP model, as regular agreement is associated with procedural memory.	and compensation by the declarative one for DLD. This would translate in a N400 ERP pattern instead of a (sustained) AN.
Consonant-final (SUBIRR) verbs used to test irregular morphosyntactic agreement	Plural condition: N400 – P600 Singular condition: Sustained AN, additional N400 and a P600.	For morphosyntactic processing of irregular forms, the DP model predicts declarative memory processingwhen accessing the stored verb form, as well as procedural memory processing for the agreement between constituents (Steinhauer and Ullman, 2002). ERP patterns found in adults suggest both declarative memory processing, illustrated by the presence of N400, and also of procedural memory system processing, signaled by the presence of sustained AN.	The PDH predicts deficits of procedural memory processing and compensation by the declarative one for DLD. This would translate in a N400 ERP pattern instead of (sustained) AN.

7. General discussion

The present thesis aimed at filling the knowledge gap regarding morphosyntactic deficits as a marker of French DLD. Subject-verb agreement difficulties, and by extension morphosyntactic deficits, have been identified as a hallmark of English-speaking preschoolers and older children with DLD. However, in studies conducted with French-speaking preschoolers with DLD, morphosyntax has not systematically been found to be a particular domain of weakness as compared to lexico-semantics. Since there is evidence that some aspects of French morphosyntax are acquired later in childhood, such as SUBIRR subject-verb number agreement, this opens the door for morphosyntax being a marker of French DLD only in older childhood and adolescence. This question was addressed in this thesis through three objectives: (1) By determining whether French-speaking (pre-)teenagers with TL have acquired SUBIRR subject-verb number agreement; (2) By resolving whether French-speaking (pre-)teenagers with DLD have impaired SUBIRR subject-verb number agreement skills when compared to teens with TL, and (3) By establishing whether SUBIRR subject-verb number agreement, and more generally morphosyntax, is a particular weakness in (pre-)teenagers with DLD when compared to other linguistic domains.

In the following sections, we will first discuss the extent to which our objectives have been met in the context of the summarised results from the three manuscripts. Next, we will review these findings in the framework of the two theories addressed in this thesis, the procedural deficit hypothesis (PDH) and the general slowing hypothesis (GSH), and extend the discussion to other studies that have explored these theories. Then, we will discuss the benefits of using neurocognitive tools such as ERPs in the study of DLD. Finally, clinical implications of this thesis are presented by providing recommendations for SLPs working with French-speaking adolescents with DLD.

7.1. It takes time to acquire subject-verb number agreement in French

Participants with no history of language impairment, aged between 7 and 14 years were included in this thesis. Their morphosyntactic development was evaluated at the behavioural level, with tasks typical of clinical and research settings, and at the neurocognitive level, with ERPs. The thesis' first objective was to determine if they have acquired SUBIRR subject-verb number agreement, as Franck et al. (2004) suggested that the very frequent irregular verb *faire* 'to do' was mastered only by age 8 in elicitation. Similar to these authors, we examined SUBIRR subject-verb number agreement using a sentence production task, but with a large sample of 20 irregular French verbs (fLEX test, Pourquié, 2015), and with older participants then those in Franck's (2004) study. Results from the first manuscript showed that participants with TL made an average of 5% of errors (i.e., 1 error in 20 verbs), a percentage that resembles the results of Franck et al. (2004) and is reminiscent of their adult data (Franck and Nicol, in preparation, cited in Franck et al., 2004). However, looking at individual score distributions of our participants, half of them actually made more than just one error. Our results suggest that a substantial proportion typically developing French-speaking children and adolescents have not yet fully mastered the production of SUBIRR verb agreement. This is not surprising, since our investigations at the neurocognitive level revealed that their ERP patterns were immature compared to adults, as shown in manuscript #2. In response to SUBIRR subject-verb number agreement mismatches in our auditory-visual bimodal paradigm (e.g., in consonant-final verb conditions hearing *ils rugissent* [ilʁyʒɪs] 'they roar' while looking at a picture of one single lion), in both singular and plural conditions, adults elicited classic biphasic patterns: a negativity followed by a positivity. The singular conditions elicited a sustained anterior negativity superimposed by a P600, and the plural ones an N400 followed by a P600. Preteenagers and teenagers exhibited, in the former condition, a late onset P600 followed by a sentence-final negativity, and in the latter condition an N400. Our results suggest that SUBIRR subject-verb number agreement is not fully acquired by French-speaking (pre-)teenagers with TL. Integrating the previous ERP literature related to morphosyntactic processing in youth with and without DLD across languages, as well as second-language learners trajectory from novice to native-like, proposed by Steinhauer et al. (2009), we elaborated in manuscript #3 (see section 5.1.2, Table 1) an interpretative grid of morphosyntactic development based on ERP patterns. At best, the TL participants exhibited an intermediate level of morphosyntactic development for SUBIRR subject-verb number agreement and did not reach an expert level of morphosyntactic processing, i.e., the adult pattern. Overall, evidence at the behavioural and the neurocognitive level suggest that French speaking (pre-)teenagers are still consolidating SUBIRR subject-verb number agreement.

7.2. Deficits on verb number agreement are a marker of French DLD in (pre-)adolescents

Teenagers with DLD between 12 and 15 years participated in our experiments. Their morphosyntactic development was evaluated at the behavioural level with tasks typical of clinical and research settings, and at the neurocognitive level with ERPs. The thesis' second objective was to resolve whether French-speaking (pre-)teenagers with DLD have impaired SUBIRR subject-verb number agreement when compared to teens with TL. On the one hand, studies have suggested that subject-verb agreement is impaired in French-speaking older children and adolescents with DLD (e.g., Franck et al., 2004; Rose & Royle, 1999). On the other hand, Elin Thordardottir et al.'s (2011) results indicated that preschoolers with DLD made more errors than children with TL in tasks assessing verb morphology, but that these tasks were not informative when discriminating between children with and without DLD. Using new robust statistical methods that are less affected by outliers and diagnostic statistics typical of clinical studies, results from manuscript #1 indicated that the task assessing SUBIRR number agreement via elicited sentence production (fLEX test, Pourquié, 2015) was one of the 3 most reliable tasks to discriminate between French-speaking teenagers with and without DLD. Our results confirm that SUBIRR subject-verb agreement production is impaired in DLD, but also that it is a reliable clinical marker for French DLD in (pre-)teenagers. At the neurocognitive level, in manuscript #3 (see section 5.1.2, Table 1), adolescents with DLD exhibited lower levels of proficiency when compared to the TL group on number agreement of SUBIRR[18] singular verb mismatches. The DLD group displayed a late frontal positivity reminiscent of a late P3a, which indexes domain-general engagement of attentional processes, and thus reflects a low to intermediate level of morphosyntactic development. The TL group exhibited a P600 which points to an intermediate developmental level. Regarding processing on number agreement of SUBIRR verb plural mismatches, both DLD and TL groups exhibited an N400, which corresponds to a low level of morphosyntactic processing. In this condition, it is not clear whether these N400s exhibited by both groups reflected the same one as the adults' biphasic N400-P600, or immature processing. This result was surprising when related to their behavioural performance, where participants with DLD made significantly more errors than those with TL in the production of SUBIRR number agreement verbs in sentence context. Nevertheless, and overall, the results suggest that deficits in

[18] These are labeled "consonant-final verbs" in manuscript #3

number agreement of SUBIRR verbs at the behavioural level are a marker for French DLD in adolescents, and differences between DLD and TL groups were observed at the neurocognitive level.

7.3. Morphosyntactic deficits are a distinctive weakness of French (pre-)adolescents with DLD

The third objective of this thesis was to establish whether morphosyntax is a particular weakness in (pre-)teenagers with DLD when compared to other linguistic domains. Specifically, we wanted to see if morphosyntax was impaired when compared to lexico-semantics, as this was suggested for preschoolers and older English-speaking children with DLD (Conti-Ramsden et al., 2001; Leonard et al., 1999), but not for French-speaking preschoolers with DLD (Elin Thordardottir et al., 2011; Elin Thordardottir & Namazi, 2007). To do so, at the behavioural level, we administered 20 subtasks to our participants assessing morphosyntax, lexico-semantics, and working memory. By focusing on developmental changes that occur between late-childhood and early adolescence, the present thesis revealed that morphosyntax stands out as a distinctive weakness in French-speaking (pre-)adolescents with DLD, based on evidence at the behavioural and neurocognitive levels. In the first manuscript, results indicated that the three most reliable tasks in discriminating between DLD and TL included a SUBIRR production task (fLEX, Pourquié, 2015), but also lexico-semantic and sentence repetition tasks (Word Classes, Recalling sentences; CELF-IV$^{cnd\text{-}F}$ Secord et al., 2009). However, when we combined our most discriminating tasks, we found that the SUBIRR verb number-agreement production task contributed the most to the multivariable model for the identification of participants with DLD. In the third manuscript, comparisons of acceptability judgements showed that the DLD group was less accurate than the TL one at identifying mismatches on morphosyntactic conditions, whereas both groups performed similarly on the lexico-semantic ones. Similarly, at the neurocognitive level, differences between groups were found on the majority of our morphosyntactic conditions–on number agreement within noun phrases, as well as on subject-verb agreement instantiated by liaison and SUBIRR verbs–while in contrast the lexico-semantic conditions elicited qualitatively similar ERP patterns in both groups.

Overall, the present thesis filled a knowledge gap regarding morphosyntactic deficits as a marker of French DLD. As outlined earlier, in studies conducted with French-speaking preschoolers with DLD, morphosyntax has not systematically been found to be a clinical marker of DLD. How can these differences between French preschool children and teenagers with DLD be explained? We think there could be two main reasons for this. First, SUBIRR subject-verb number agreement skills are under-informative of language acquisition in preschool children as they are still in development. Second, we think the tasks used to evaluate morphosyntax in children play an important role. The two studies that compared morphosyntax to lexico-semantics in preschoolers used either only spontaneous speech (Elin Thordardottir & Namazi, 2007) or a task that assesses comprehension of multiple instantiations of (morpho-)syntax: subject-verb number and tense agreement as well as syntax (Carrow/TALC; Elin Thordardottir et al., 2011). These tasks probably did not provide contexts in which they could target deficits specific to DLD in French. Indeed, Rvachew et al. (2017) showed that production of past-tense *passé composé* was a reliable marker to identify school-aged children, including 6-year-olds, with language difficulties. These authors developed a screener for literacy delay in French-speaking first-graders (Phophlo; *ibid*) which contains 10 verbs taken from a French past-tense verb production task (Royle et al., 2018). While this screener has low sensitivity (71%), it shows excellent specificity (94%) for the identification of children who will eventually develop spelling difficulties. It also identified children who had or were suspected of having DLD and were awaiting evaluation. Verb inflection production, phoneme production and rhyme awareness were all found to be important predictors of oral and written language disabilities. The use of this type of task, precisely past-tense production, in studies comparing linguistic abilities–i.e., morphosyntax to lexico-semantics–of French-speaking 5- or 6-year-old children with DLD–could reveal that morphosyntax is a particular weakness in younger French-speaking children with DLD.

7.4. Two neurocognitive theories of DLD: The procedural deficit hypothesis and the general slowing hypothesis

We will now turn to the two theories addressed in this thesis, the PDH and the GSH. We will review our findings within the two accounts' respective frameworks and extend the discussion to other studies that have explored these theories.

The PDH predicts that language structures that are underpinned by the declarative memory system are processed the same way by our participants with and without DLD, whereas the ones underpinned by the procedural memory system are processed differently, as it is expected to be impaired. The first manuscript's results showed that the three most discriminant tasks were Recalling Sentences and Word Classes from the CELF-IV^{cnd-F} (Secord et al., 2009) and production of SUBIRR subject-verb number agreement (fLEX, Pourquié, 2015). Deficits in these first two tasks are in line with the PDH predictions. Sentence repetition relies on phonological working memory, as well as (morpho-)syntactic and lexico-semantic skills (Leclercq et al., 2014), the first two domains being strongly associated with the procedural system. Word Classes is a multi-word task that evaluates lexico-semantics. Participants were asked to compare four spoken words and identify the two that go together based on a semantic relationship. It is coherent with the PDH that this task was especially hard for teenagers with DLD. The procedural system underlies speech perception and word segmentation, which are used in word learning and thus for developing semantic representations (Ullman et al., 2020). DLD would be associated with weaker semantic representations of known words, and comparing words based on meaning will tap into this deficit. Consistent with the PDH, the one-word vocabulary task EVIP (the French PPVT, Dunn et al., 1993) was less descriminating between the DLD and TL groups. Indeed, this type of task assessing lexical receptive skills would be associated with the declarative system. The third most discriminating task between our groups was production of SUBIRR subject-verb number agreement, and most errors occurred in plural sentences, where DLD participants produced the singular verb form (Pourquié et al., in revision). This result is partly consistent with the PDH in that it assumes that irregular word forms are more easily processed by the declarative system because they can be chunked at the word level, so this task should not be one of the most discriminating between our groups. However, the sentence-level morphosyntactic checking that this task demands relies on the procedural system and is thus expected to be impaired (Steinhauer & Ullman, 2002).

In sum, results from the first manuscript are globally in line with the PDH, as were the findings in the third manuscript. Results indicated that lexico-semantic acceptability judgements and ERPs were similar between groups, while several differences were found on morphosyntactic conditions (see discussion in manuscript #3, section 5.4.3, for details). In the present thesis, we tested the predictions of the PDH on linguistic tasks as well as on linguistic aspects of sentence processing that were assigned to either the declarative or the procedural system, and found that these predictions were in general accurate. However, studies that have examined the PDH in children with DLD have typically used tasks designed to assess participants' ability to learn sequences, either explicitly by tapping into declarative memory, or implicitly by tapping into the procedural one, and they have found mixed results. Multiple studies found links between grammar skills and implicit sequence learning, suggesting that (morpho-)syntactic deficits in DLD are associated with poor procedural learning (Conti-Ramsden et al., 2015; Hedenius et al., 2011; Kuppuraj & Prema, 2013). However, others did not find evidence supporting the PDH. Bani Hani (Bani Hani, 2015; Bani Hani and Nadig, 2014) observed similar performance in children with and without DLD on a serial reaction time task. Poll et al. (2015) tested performance in two groups of adults, with and without DLD, on tasks assessing either the procedural or declarative memory system. They found that the DLD group's performance on both tasks was worse than that of the control group, and that the effect size on the declarative task scores was *larger* than on the procedural task. The authors point out that this contradicts the predictions of the PDH: individuals with DLD who are at a later stage of development should have an intact or better declarative memory capacity than the TL group. Gabriel et al. (2015) found that among French-speaking children aged 7-13 years with and without DLD, the DLD group performed worse than the control group on an auditory task assessing procedural memory, while they performed similarly well on the same task with visual stimuli. What could explain these mixed results? First, we can question the psychometric properties of tasks evaluating the procedural and declarative memory system. The tasks used in these studies are generally experimental and have not been standardised or modelled to ensure consistency between the feature of interest, i.e., the two memory systems, and the task results, which is important when measuring a psychological attribute (Briggs, 2022). West et al., (2021) recently published a meta-analytic review of studies that used procedural memory tasks with groups of individuals with DLD or dyslexia and found that these groups performed more poorly on all procedural tasks. They also analysed correlational studies of procedural memory and language tasks in studies of TD children. West et al.'s (2021) pooled effect size suggests no correlation between language abilities and performance on the serial reaction time task. Second, these studies tested the PDH predictions only on task outcomes. Eisenberg et al. (2019), suggest that theories that predict behaviour in experimental contexts only may have no relevance to natural behavioural outcomes, and that the constructed nature of experimental tasks may fundamentally affect their ecological validity. We think that one strength of the PDH is that predictions can be made not only regarding tasks in experimental contexts but also on everyday activities such as language. These mixed results were found between studies testing PDH predictions on experimental tasks tapping into the procedural and declarative systems. All ERP and most behavioural results of the present dissertation generally support the PDH based on grammatical sentence processing, reminiscent of more naturalistic contexts. Future studies testing predictions of the PDH and more generally DLD should use a combination of experimental tasks and ecologically valid activities to ensure that cognitive measures reflect real-world behaviour (Eisenberg et al., 2019).

Now turning to the second account of DLD, the GSH (Kail, 1994), predictions were tested in the third manuscript only, based on lexico-semantic and sentence processing during the ERP experiment. The GSH predictions were that, since both groups use similar processing approaches for all language tasks including morphosyntactic ones, we should observe processing speed deficits across the board in the DLD group, instantiated by longer response times and delays in ERP components compared to the TL group. We found later onset of ERP components for mismatches on the lexico-semantic condition and on determiner number agreement in noun phrases at sentence onset. Other conditions elicited either different ERP components between groups or no difference at all (see discussion in manuscript #3, section 5.4.3, for details). As outlined by Balilah et al. (2019), accounts of DLD that suggest limitations in overall capacity, such as the GSH with its limited processing speed, face the major challenge explaining disproportionate language deficits in comparison to non-linguistic domains. It thus fails both in explaining the dominance of language difficulties but also, based on the results of this dissertation and previous studies, to explain the differences observed between the different language domains. Instead of considering teens with DLD as having deficient processing speed, i.e., the cognitive function as defined by Kail and Salthouse (1994), we propose that delays observed in processing speed should be considered a consequence of impaired processes underlying language. This would be coherent with research on developmental dyslexia, where slower reading speed is observed as a result of impaired reading processes (Ziegler et al., 2008). As we found later component onsets for one of the morphosyntactic conditions and the lexico-semantic condition, these would imply impaired processes for both of these linguistic domains. This could be in contradiction with the present thesis supporting a relative strength for lexico-semantics in comparison to morphosyntax. However, we would like to emphasise that this relative strength does *not* indicate that in different lexico-semantic tasks or real language situations youth with DLD will not demonstrate linguistic deficits. Since there is evidence that people with DLD have unusual cortical and subcortical brain structures (for a review see Krishnan et al., 2016), it is to be expected that their behaviour will differ to some extent from that of their peers with TL, even in the linguistic domains that they master well.

Our results suggest that the PDH predictions are more accurate in describing language processing in DLD then the GSH. By comparing these two hypotheses, we gained a better understanding of the neurocognitive underpinnings of DLD. In the next section, we will briefly discuss why research on DLD should focus more on the neurocognition of language.

7.5. Neurocognition of language in the study of DLD

As pointed out in the introduction of this thesis, finding a defining characteristic of developmental language disorder is not an easy task, as children around the world learn different languages. Neurocognitive investigations using ERPs have the distinct advantage of allowing us to compare language processing abilities of people with DLD across different languages, as it provides us with a common basis for comparing lexical-semantic and morphosyntactic processing. When looking at ERP studies of DLD across languages, morphosyntax has been found to be a serious candidate as a defining characteristic of DLD. On the one hand, ERP studies of lexico-semantic processing have shown that children and teenagers with DLD consistently exhibit the N400 component, similar to control groups, in studies conducted in English, Dutch and Russian. French can now be added to this list. On the other hand, previous ERP studies investigating morphosyntax, mainly in English but also in Italian, show that the groups with DLD did not achieve adult-like expert morphosyntactic processing, as highlighted in our review of morphosyntactic development based on ERP patterns (see section 5.1.2, Table 1). One challenge in interpreting these results is that for most of the studies, control groups of children and adolescents with TL do not present the expected adult patterns (often biphasic LAN + P600 components), as in our ERP experiment, presented in the third manuscript (see section 6). To establish the absence of typical patterns during morphosyntactic processing as a defining feature of DLD, studies of adults with DLD also need to be conducted.

Another advantage in using ERPs in establishing a defining characteristic of DLD is that we can compare children with DLD's neurocognitive patterns to those of other groups with neurodevelopmental disorders. While the N400 has been consistently observed in response to lexical-semantic mismatches in children and adolescents with DLD, there is mixed evidence regarding children with autistic spectrum disorder (ASD). Manfredi et al. (2020) presented (pre-)adolescents with and without ASD aged 9 to 16 years with short auditory sentences with an incongruent last word. They found a typical broadly distributed N400 in centro-parietal regions for the control group, but a reduced N400 limited to the occipital regions was found for the ASD group. DiStefano et al. (2019) presented 5–11-years-olds with and without ASD a picture paired with a congruent or incongruent sound. While they found an N400 for all participants, the clinical group exhibited a later component onset with an offset characterised by a frontal distribution. Future studies should use ERPs to compare DLD with other neurodevelopmental disorders such as ASD to better understand their lexico-semantic and morphosyntactic processing. Our innovative paradigm developed in manuscript #2 would be appropriate for these neurodevelopmental populations, as grammatical sentences paired with images are similar to common language-based activities.

7.6. Limitations and future directions

We will discuss a number of potential limitations concerning the findings of the present dissertation. A first potential limitation concerns the first objective of this thesis, which was to determine if French-speaking (pre-)teenagers with TL have acquired SUBIRR subject-verb number agreement. In the present thesis, both participants with TL and adults participated in the same ERP experiment and we were able to compare them on their neurocognitive processing of SUBIRR number agreement. However, we didn't administer the behavioural task evaluating SUBIRR number agreement to the adult group (fLEX, Pourquié et al., 2015). Looking at individual data, we found that half of the (pre-)teen TL participants made more than 5% errors (i.e., 2/20 verbs or more) and we assumed that these results suggest that our TL participants had not fully acquired SUBIRR subject-verb number agreement. Since we did not have the adult results as a comparison, it could be that this profile actually reflects what should be expected of adults. Furthermore, as the TL group results showed that SUBIRR number agreement is *almost* fully acquired, it would have been interesting to test a group of younger children, for instance 7 to 10 years, where agreement and other morphosyntactic processes emerge and consolidate in TD kids. We could compare their response patterns with those of our DLD group to see if the number of errors and the verbs on which the errors occur are different or similar. A similar pattern between a younger TL group and our 14-year-old DLD group would be consistent with Paradis and Crago's (2001) "extended optional default" account, in which children with DLD are thought to have similar overuse of finite verb stems (singular present tense) as children with (TL), but over a longer period of time.

A second limitation concerns our third objective which was to establish whether morphosyntax was a particular weakness of DLD. To do so, we investigated two types of subject-verb number agreement, and number agreement on noun phrase determiners. Future studies should focus on past-tense production tasks, as they could provide even more evidence that morphosyntax is a clinical marker in French DLD. Indeed, studies have pointed to deficits in past-tense production or insensitivity to verb conjugation groups in French-speaking children and adults with DLD (Rose & Royle, 1999; Royle et al., 2018; Royle & Thordardottir, 2008).

A third limitation concerns the evaluation of the two accounts discussed in this thesis, the PDH and the GSH. These two accounts have predictions for language processing. However, they also have predictions for non-linguistic processing, and these have not been explored. To test the PDH hypotheses, we could have included tasks that assessed declarative and procedural memory more directly. Furthermore, it would have been interesting to compare non-linguistic and linguistic domains. Regarding the GSH theory, participants' processing speed was assessed at the neurocognitive level with ERP component onset, and we observed onset delays. We found no indicator of deficits in processing speed at the behavioural level. This lack of difference is most likely related to our experimental design, which included a buffer of 1000 ms between sentence offset and accuracy judgement, in order to reduce movement artefacts in the ERP signal. We should have included at least one other, crucially non-linguistic, behavioural task where reaction times could have been measured.

A fourth limitation concerns our grouping of subregular and irregular verbs from the second and third conjugation groups for the study of non-regular present tense subject-verb number agreement (see condition with Consonant-final (SUBIRR) verbs in manuscript #2 and #3). We combined these two conjugation groups because they both have different phonological forms in the production of singular and plural present tense. In doing this, we have assumed that production of verbs in the present tense from both groups would be 'stored' as the verb stem is changing and the verb forms cannot be 'derived from a rule'. It should be noted that this is not so simple when considering studies of past tense verb comprehension and production. On the one hand, as summarised in a study by Royle, Beritognolo and Bergeron, (2012) on regularity, subregularity and irregularity, Kresh (2008) found that in a lexical decision task on past participle forms, French-speaking children and adults had faster reaction times and lower error rates for the first and second conjugation groups. Results of lexical decision on verbs from the third group were influenced by verb frequency: longer reaction times and higher error rates were found on lower frequency form than for high-frequency ones. Based on this study, we could conclude that only the third conjugation group verb forms are stored in the lexicon, as it was the only conjugation group subject to verb frequency effect. However, on the other hand, Royle, Beritognolo and Bergeron (2012) commented that for verb production, the story is more complex. During the study of production of pas tense verbs in French-speaking children, Royle 2007 found that young children were more proficient with verbs from the first (regular) and second (subregular) conjugation groups than the third group's (irregular) verbs. Nevertheless, results also revealed a frequency effect, i.e., low-frequency forms were not produced well, mainly when they did not belong into the first conjugation regular pattern. Based on these results, we could conclude that the second and third conjugation groups are stored because there were influenced by frequency effect. Overall, based on these results, it is not clear if second group subregular verbs are more similar to the regular verbs from the first conjugation groups or the irregular verbs from the third conjugation group. See Royle, Beritognolo and Bergeron (2012) for further discussion on this topic within a psycholinguistic framework. To address this debate within the scope of this thesis, we could have done ERP analyses based on items and see if the subregular verbs ERP patterns differ from the irregular ones in the condition regrouping them, i.e., within the SUBIRR consonant-final verb.

We believe that the main obstacle to confirming morphosyntax as a defining feature of DLD is the lack of studies studying morphosyntactic processing at the behavioural and neurocognitive level in (young) adults with DLD. Our results indicated that teenagers with DLD have deficits in production and processing of SUBIRR subject-verb agreement. Our results and previous studies have shown that teenagers with DLD do not exhibit adult-like morphosyntactic processing at the neurocognitive level. However, there is always the possibility that they will reach mature processing later in early adulthood. Studies investigating adolescents have sometimes included young adults in their participants (e.g., Weber-Fox et al., 2010, tested a group of teenagers with DLD aged 14;3 to 18;1 years, with an average of 15;9), but to our knowledge, no ERP study has addressed morphosyntactic processing in a DLD population aged 18 years old and above.

7.7. Conclusion and clinical recommendations based on thesis results

Through three objectives, the present thesis filled a knowledge gap concerning morphosyntactic deficits as a marker of French DLD. Our findings strongly imply that SUBIRR subject-verb number agreement deficits are a reliable marker of DLD in French teenagers, and that morphosyntax is a distinct area of weakness for them when compared to lexico-semantics. From the results of this thesis, recommendations can be derived for SLPs working with French-speaking adolescents with DLD. We present our recommendations in a way that they can be easily integrated into clinical practice, as does the popular SLP blog *Tout cuit dans le bec* (Di Sante and Gingras, n.d., roughly meaning "ready to eat" or "straightforward and easy to understand") . The main objective of these recommendations is to support evidence-based practice. This list does not claim to be exhaustive, but addresses the most important findings of this dissertation.

- Deficits in the production of subregular and irregular subject-verb number agreement are a clinical marker of DLD in French-speaking adolescents. Note that it is, however, expected that typically developing French-speaking (pre-)teenagers from 7 to 15 years-old could normally produce 1 or 2 errors out of a list of 20 verbs. While no standard task is currently available, the list of 20 verbs used in this thesis will be available in Pourquié et al. (in revision).
- Deficits on the CELF-IV$^{cnd\text{-}F}$ Word Classes task assessing receptive lexico-semantic skills (Secord et al., 2009), are a reliable marker for French-speaking adolescents with DLD.

- Deficits on the Recalling Sentences CELF-IV[cnd-F] task (Secord et al., 2009), are a reliable marker for French-speaking adolescents with DLD.
- Clinicians using the CELF-IV[cnd-F] (Secord et al., 2009) Word Classes and Recalling Sentences tasks should be careful when using the French-Canadian norms, as our results suggest that optimal cut-offs scores between teenagers with and without DLD are far above the recommended 16th percentile. We suggest that a 25th percentile score on these tasks should be used to identify whether a teenager from Québec is impaired.
- When listening to a sentence, French-speaking adolescents with DLD have difficulty identifying subject number, either plural or singular, based on the verb form only. Adding more explicit and lexical number cues will help them identify subject number (e.g., *Ils dorment* 'they sleep' is more difficult to understand than *Les deux garçons dorment* 'the two boys sleep', where both the determiner and the numeral carry number information).
- Although instruction occurs primarily through written language in high school, it is essential that intervention for adolescents with DLD continues to target oral language, as this remains the source of their difficulties whether in the oral or written form.

7.8. References

Balilah, A., Rafat, Y., & Archibald, L. (2019). Domain-Specific and Domain-General Processing Accounts in Children with Specific Language Impairment (SLI): Contribution of Cross-Linguistic Evidence. In S. Hidri (Ed.), *English Language Teaching Research in the Middle East and North Africa: Multiple Perspectives* (pp. 383–407). Springer International Publishing. https://doi.org/10.1007/978-3-319-98533-6_18

Bani Hani, H. (2015). *Language-impaired children with autism spectrum disorders and children with specific language impairment: Similar language abilities but distinct memory profiles* [Thèse de doctorat, McGill]. http://digitool.library.mcgill.ca/R/5NR8DC5BRI2CTCREAE6QKIATJY9Q3SC36A92J6UXGXTU2EEG5B-05764

Bani Hani, H., & Nadig, A. (2014). *Autism with Language Impairment vs. Specific Language Impairment: Different Declarative and Procedural Memory Profiles.* American Speech-Language-Hearing Association Conference, Orlando, Floride. http://poplab.mcgill.ca/pubmanager/uploads/BANIHANI-NADIG-ASHA2014_6558732328.pdf

Briggs, D. C. (2022). What is Measurement? In *Historical and Conceptual Foundations of Measurement in the Human Sciences: Credos and Controversies* (pp. 1–24). Routledge. https://doi.org/10.1201/9780429275326

Conti-Ramsden, G., Botting, N., & Faragher, B. (2001). Psycholinguistic markers for specific language impairment (SLI). *Journal of Child Psychology and Psychiatry*, *42*(6), 741–748. https://doi.org/10.1111/1469-7610.00770

Conti-Ramsden, G., Ullman, M. T., & Lum, J. A. G. (2015). The relation between receptive grammar and procedural, declarative, and working memory in specific language impairment. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01090

Di Sante, M., & Gingras, M.-P. (n.d.). *Tout cuit dans le bec*. Tout cuit dans le bec. Retrieved February 11, 2022, from https://cuitdanslebec.wordpress.com/

DiStefano, C., Senturk, D., & Jeste, S. S. (2019). ERP evidence of semantic processing in children with ASD. *Developmental Cognitive Neuroscience*, *36*, 100640. https://doi.org/10.1016/j.dcn.2019.100640

Eisenberg, I. W., Bissett, P. G., Zeynep Enkavi, A., Li, J., MacKinnon, D. P., Marsch, L. A., & Poldrack, R. A. (2019). Uncovering the structure of self-regulation through data-driven ontology discovery. *Nature Communications*, *10*(1), 2319. https://doi.org/10.1038/s41467-019-10301-1

Elin Thordardottir, Kehayia, E., Mazer, B., Lessard, N., Majnemer, A., Sutton, A., Trudeau, N., & Chilingaryan, G. (2011). Sensitivity and specificity of French language and processing measures for the identification of primary language impairment at age 5. *Journal of Speech, Language, and Hearing Research*, *54*(2), 580–597. https://doi.org/10.1044/1092-4388(2010/09-0196)

Elin Thordardottir, & Namazi, M. (2007). Specific language impairment in French-speaking children: Beyond grammatical morphology. *Journal of Speech, Language, and Hearing Research*, *50*(3), 698–715. https://doi.org/1092-4388/07/5003-0698

Franck, J., Cronel-Ohayon, S., Chillier, L., Frauenfelder, U. H., Hamann, C., Rizzi, L., & Zesiger, P. (2004). Normal and pathological development of subject–verb agreement in speech production: A study on French children. *Journal of Neurolinguistics*, *17*(2), 147–180. https://doi.org/10.1016/S0911-6044(03)00057-5

Gabriel, A., Meulemans, T., Parisse, C., & Maillart, C. (2015). Procedural learning across modalities in French-speaking children with specific language impairment. *Applied Psycholinguistics*, *36*(3), 747–769. https://doi.org/10.1017/S0142716413000490

Hedenius, M., Persson, J., Tremblay, A., Adi-Japha, E., Veríssimo, J., Dye, C. D., Alm, P., Jennische, M., Tomblin, J. B., & Ullman, M. T. (2011). Grammar predicts procedural learning and consolidation deficits in children with specific language impairment. Research in Developmental Disabilities, 32(6), 2362–2375. https://doi.org/10.1016/j.ridd.2011.07.026

Kresh, S. (2008). L'acquisition et le traitement de la morphologie du participe passé du français (Doctoral dissertation, Université du Québec à Montréal).

Krishnan, S., Watkins, K. E., & Bishop, D. V. M. (2016). Neurobiological Basis of Language Learning Difficulties. *Trends in Cognitive Sciences*, *20*(9), 701–714. https://doi.org/10.1016/j.tics.2016.06.012

Kuppuraj, S., & Prema, K. S. R. (2013). Aspects of grammar sensitive to procedural memory deficits in children with specific language impairment. *Research in Developmental Disabilities*, *34*(10), 3317–3331. https://doi.org/10.1016/j.ridd.2013.06.036

Leclercq, A.-L., Quémart, P., Magis, D., & Maillart, C. (2014). The sentence repetition task: A powerful diagnostic tool for French children with specific language impairment. *Research in Developmental Disabilities*, *35*(12), 3423–3430. https://doi.org/10.1016/j.ridd.2014.08.026

Leonard, L. B., Miller, C., & Gerber, E. (1999). Grammatical morphology and the lexicon in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, *42*(3), 678–689. https://doi.org/10.1044/jslhr.4203.678

Manfredi, M., Cohn, N., Sanchez Mello, P., Fernandez, E., & Boggio, P. S. (2020). Visual and Verbal Narrative Comprehension in Children and Adolescents with Autism Spectrum

Disorders: An ERP Study. *Journal of Autism and Developmental Disorders*, *50*(8), 2658–2672. https://doi.org/10.1007/s10803-020-04374-x

Poll, G. H., Miller, C. A., & van Hell, J. G. (2015). Evidence of compensatory processing in adults with developmental language impairment: Testing the predictions of the procedural deficit hypothesis. *Journal of Communication Disorders*, *53*, 84–102. https://doi.org/10.1016/j.jcomdis.2015.01.004

Pourquié, M. (2015). fLEX: Multilingual assessment of inFlectional and LEXical processing. *Software, Intellectual Property 2016-01*, *88*.

Pourquié, M., Courteau, É., Duquette, A.-S., & Royle, P. (in revision). *Verb inflection and argument structure processing in French adolescents with DLD.*

Rose, Y., & Royle, P. (1999). Uninflected structure in familial language impairment: Evidence from French. Folia Phoniatrica et Logopaedica, 51(1–2), 70–90. https://doi.org/10.1159/000021482

Royle, P. (2007). Variable effects of morphology and frequency on inflection patterns in French preschoolers. *The Mental Lexicon*, *2*(1), 103–125.

Royle, P., Beritognolo, G., & Bergeron, E. (2012). Regularity, sub-regularity and irregularity in French acquisition. Irregularity in morphology (and beyond). Berlin: Akademie Verlag, 227-250.

Royle, P., St-Denis, A., Mazzocca, P., & Marquis, A. (2018). Insensitivity to verb conjugation patterns in French children with SLI. *Clinical Linguistics & Phonetics*, *32*(2), 128–147. https://doi.org/10.1080/02699206.2017.1328706

Royle, P., & Thordardottir, E. T. (2008). Elicitation of the passé composé in French preschoolers with and without specific language impairment. *Applied Psycholinguistics*, *29*(3), 341–365. https://doi.org/10.1017/S0142716408080168

Rvachew, S., Royle, P., Gonnerman, L. M., Stanké, B., Marquis, A., & Herbay, A. (2017). Development of a tool to screen risk of literacy delays in French-speaking children: PHOPHLO. *Canadian Journal of Speech Language Pathology and Audiology = Revue Canadienne d'orthophonie et d'audiologie*, *41*, 321–340.

Secord, W. A., Wiig, E., Boulianne, L., Semel, E., & Labelle, M. (2009). *Évaluation clinique des notions langagières fondamentales®—Version pour francophones du Canada (CELF® CDN-F)*. The Psychological Corporation.

Steinhauer, K., & Ullman, M. T. (2002). Consecutive ERP effects of morpho-phonology and morpho-syntax. *Brain and Language*, *83*, 62–65.

Steinhauer, K., White, E. J., & Drury, J. E. (2009). Temporal dynamics of late second language acquisition: Evidence from event-related brain potentials. *Second Language Research*, *25*(1), 13–41. https://doi.org/10.1177/0267658308098995

Ullman, M. T., Earle, F. S., Walenski, M., & Janacsek, K. (2020). The Neurocognition of Developmental Disorders of Language. *Annual Review of Psychology*, *71*(1), null. https://doi.org/10.1146/annurev-psych-122216-011555

Weber-Fox, C., Leonard, L. B., Wray, A. H., & Tomblin, J. B. (2010). Electrophysiological correlates of rapid auditory and linguistic processing in adolescents with specific language impairment. *Brain and Language*, *115*(3), 162–181. https://doi.org/10.1016/j.bandl.2010.09.001

West, G., Melby-Lervåg, M., & Hulme, C. (2021). Is a procedural learning deficit a causal risk factor for developmental language disorder or dyslexia? A meta-analytic review. *Developmental Psychology*, *57*(5), 749–770. https://doi.org/10.1037/dev0001172

Ziegler, J. C., Castel, C., Pech-Georgel, C., George, F., Alario, F.-X., & Perry, C. (2008). Developmental dyslexia and the dual route model of reading: Simulating individual differences and subtypes. *Cognition*, *107*(1), 151–178. https://doi.org/10.1016/j.cognition.2007.09.004