



BRAIN-BEHAVIOUR INTERFACES IN LINGUISTIC COMMUNICATION

EDITED BY: Yury Y. Shtyrov, Andriy Myachykov, Beatriz Martín-Luengo and
Olga V. Shcherbakova

PUBLISHED IN: *Frontiers in Human Neuroscience*, *Frontiers in Communication* and
Frontiers in Psychology



frontiers

Frontiers eBook Copyright Statement

The copyright in the text of individual articles in this eBook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this eBook is the property of Frontiers.

Each article within this eBook, and the eBook itself, are published under the most recent version of the Creative Commons CC-BY licence.

The version current at the date of publication of this eBook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or eBook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714
ISBN 978-2-88966-142-8
DOI 10.3389/978-2-88966-142-8

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

BRAIN-BEHAVIOUR INTERFACES IN LINGUISTIC COMMUNICATION

Topic Editors:

Yury Y. Shtyrov, Aarhus University, Denmark

Andriy Myachykov, Northumbria University, United Kingdom

Beatriz Martín-Luengo, National Research University Higher School of Economics, Russia

Olga V. Shcherbakova, Saint Petersburg State University, Russia

Citation: Shtyrov, Y. Y., Myachykov, A., Martín-Luengo, B., Shcherbakova, O. V., eds. (2020). Brain-Behaviour Interfaces in Linguistic Communication. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88966-142-8

Table of Contents

- 05** ***Editorial: Brain-Behaviour Interfaces in Linguistic Communication***
Olga Shcherbakova, Andriy Myachykov, Beatriz Martín-Luengo and Yury Shtyrov
- 07** ***Effects of Self-Enhancement on Eye Movements During Reading***
Ya Lou, Huajian Cai, Xuewei Liu and Xingshan Li
- 12** ***Embodied Semantics in a Second Language: Critical Review and Clinical Implications***
Elisa Monaco, Lea B. Jost, Pascal M. Gygax and Jean-Marie Annoni
- 33** ***Female Speakers Benefit More Than Male Speakers From Prosodic Charisma Training—A Before-After Analysis of 12-Weeks and 4-h Courses***
Oliver Niebuhr, Silke Tegtmeier and Tim Schweisfurth
- 39** ***Explicitly Slow, Implicitly Fast, or the Other Way Around? Brain Mechanisms for Word Acquisition***
Yury Shtyrov, Alexander Kirsanov and Olga Shcherbakova
- 43** ***Semantic Processing in Bilingual Aphasia: Evidence of Language Dependency***
Marco Calabria, Nicholas Grunden, Mariona Serra, Carmen García-Sánchez and Albert Costa
- 58** ***Concrete vs. Abstract Semantics: From Mental Representations to Functional Brain Mapping***
Nadezhda Mkrtychian, Evgeny Blagovechtchenski, Diana Kurmakaeva, Daria Gnedykh, Svetlana Kostromina and Yury Shtyrov
- 64** ***Effects of Visual Priming and Event Orientation on Word Order Choice in Russian Sentence Production***
Mikhail Pokhoday, Yury Shtyrov and Andriy Myachykov
- 72** ***Speed-Accuracy Tradeoffs in Brain and Behavior: Testing the Independence of P300 and N400 Related Processes in Behavioral Responses to Sentence Categorization***
Phillip M. Alday and Franziska Kretzschmar
- 94** ***Identifying the Speech Production Stages in Early and Late Adulthood by Using Electroencephalography***
Jakolien den Hollander, Roel Jonkers, Peter Mariën and Roelien Bastiaanse
- 115** ***Neurophysiological Correlates of Fast Mapping of Novel Words in the Adult Brain***
Marina J. Vasilyeva, Veronika M. Knyazeva, Aleksander A. Aleksandrov and Yury Shtyrov
- 125** ***Auditory Mismatch Negativity Response in Institutionalized Children***
Irina Ovchinnikova, Marina A. Zhukova, Anna Luchina, Maxim V. Petrov, Marina J. Vasilyeva and Elena L. Grigorenko
- 133** ***Novel Word Learning: Event-Related Brain Potentials Reflect Pure Lexical and Task-Related Effects***
Beatriz Bermúdez-Margaretto, David Beltrán, Fernando Cuetos and Alberto Domínguez

146 *Language Processing as a Precursor to Language Change: Evidence From Icelandic*

Ina Bornkessel-Schlesewsky, Dietmar Roehm, Robert Mailhammer and Matthias Schlewsky

164 *An Eye-Tracking Study of Sketch Processing: Evidence From Russian*

Tatiana E. Petrova, Elena I. Riekhakaynen and Valentina S. Bratash



Editorial: Brain-Behaviour Interfaces in Linguistic Communication

Olga Shcherbakova^{1,2*}, Andriy Myachykov^{3,4}, Beatriz Martín-Luengo⁴ and Yury Shtyrov^{2,5}

¹ Department of General Psychology, Faculty of Psychology, Saint Petersburg State University, Saint Petersburg, Russia, ² Laboratory of Behavioural Neurodynamics, Saint Petersburg State University, Saint Petersburg, Russia, ³ Department of Psychology, Northumbria University, Newcastle upon Tyne, United Kingdom, ⁴ Centre for Cognition and Decision Making, National Research University Higher School of Economics, Moscow, Russia, ⁵ Department of Clinical Medicine, Center of Functionally Integrative Neuroscience, Aarhus University, Aarhus, Denmark

Keywords: brain, language, EEG, neurolinguistic, psycholinguistic, embodied cognition

Editorial on the Research Topic

Brain-Behaviour Interfaces in Linguistic Communication

Language is a uniquely human cognitive function, which greatly defines and determines our psychological and social traits. Despite the importance of language and speech, they remain among the least understood human cognitive processes, and their neurobiological underpinnings are still poorly understood.

In recent decades, an immense body of diverse data illuminating the neural bases of language processes in both children and adults has been acquired through the use of many advanced techniques. These include electroencephalography (EEG), magnetoencephalography (MEG), functional magnetic-resonance imaging (fMRI), transcranial magnetic stimulation (TMS), transcranial direct and alternating current stimulation (tDCS, tACS), eye-tracking, behavioral measures, etc. The combined power of these techniques continues to shed light upon the brain mechanisms of language acquisition, comprehension and processing, speech disorders, their diagnosis and treatment, as well as the interplay between language and other neurocognitive systems and functions.

The aim of the Research Topic *Brain-Behavior Interfaces in Linguistic Communication* is to provide a state-of-the-art overview of this diverse and multidisciplinary area of research, with a special emphasis on bridging the gap between different research fields, theoretical views, and methodologies.

Our Research Topic offers a collection of 14 articles on various facets of linguistic behavior and its neural underpinnings. The collection comprises 11 research papers (including six original research reports and five brief research reports), one comprehensive review, one mini review, and one opinion paper.

The collection can be topically divided into several groups of papers. The first group brings together several articles using electroencephalography in order to investigate *the neural bases* of language learning and use. The opinion article by Shtyrov et al. addresses the effectiveness and neural underpinning of two main routes of *novel word acquisition*: (1) explicit encoding and (2) implicit learning (fast mapping). The authors discuss methodological confounds besetting existing research paradigms and provide a clear perspective for designing a comprehensive and fully balanced experimental approach for comparing these two language learning modes. The experimental study described by Vasilyeva et al. follows up on this and investigates the neural bases of *fast mapping* in adults by documenting near-instant changes in neural activity after a single-shot novel word training. The authors conclude that fast mapping may promote rapid integration of newly learned items into the brain's neural lexicon, even into adulthood. In a related article on *ERP correlates of novel word learning*, Bermúdez-Margaretto et al. show how novel words repeatedly associated

OPEN ACCESS

Edited and reviewed by:

Kirrie J. Ballard,
The University of Sydney, Australia

*Correspondence:

Olga Shcherbakova
o.shcherbakova@spbu.ru

Specialty section:

This article was submitted to
Speech and Language,
a section of the journal
Frontiers in Human Neuroscience

Received: 23 June 2020

Accepted: 21 July 2020

Published: 11 September 2020

Citation:

Shcherbakova O, Myachykov A,
Martín-Luengo B and Shtyrov Y (2020)
Editorial: Brain-Behaviour Interfaces in
Linguistic Communication.
Front. Hum. Neurosci. 14:324.
doi: 10.3389/fnhum.2020.00324

with meaningful cues demonstrate a higher attenuation of N400 responses than the words trained in a basic orthographic condition, confirming facilitation of the lexico-semantic processing of these stimuli as a consequence of semantic association. This finding suggests that novel word learning could be influenced by the activation of the categorization-related network. Next, the contribution by Ovchinnikova et al. investigated auditory event-related potentials in children reared in two very different types of environment: biological-family care or institutional care. The paper makes an important contribution concerning *the role of social environment in neurocognitive maturation*. den Hollander et al. further inform this debate by using EEG for identifying *the speech production stages* in early and late adulthood. They report no scalp distribution differences between the two groups suggesting that the same networks are involved at different stages, regardless of the age, even though the timing of the individual stages is different between the groups. Alday and Kretzschmar used ERP and multiple-response speed-accuracy trade-off (SAT) paradigm to investigate *the relationship between N400 and P300 ERP components*. The article clarifies how these two classic ERP potentials determine behavioral profiles. With the use of multivariate Bayesian mixed-effects models, GLMM-based approach, and partial effects, the paper demonstrates how overlapping ERP responses in one sample of participants predict behavioral SAT profiles of another sample. Moreover, this research confirms that the P300 and N400 reflect two independent but interacting processes and that the competition between these processes is reflected differently in the speed-accuracy trade-off behavior. Finally, in an EEG study on *a language in transition* (Icelandic) Bornkessel-Schlesewsky et al. show that the neurophysiological responses already reflect projected language changes that are not yet apparent in the overt behavior of native speakers.

Another set of articles address *semantic aspects* of language learning and use. The mini review by Mkrtychian et al. offers a snapshot of psycholinguistic and neurocognitive approaches to studying *concrete and abstract semantics*. A review by Monaco et al. discusses the role of *embodied semantics in second language comprehension* arguing that L2 is embodied differently than L1 (which might have important clinical implications). Lastly, the research by Calabria et al. addresses the issue of *semantic processing in bilingual* (Catalan—Spanish) *aphasia*. The results suggest that lexical retrieval in individuals with bilingual aphasia may be selectively impaired within their non-dominant language due to an excessive amount of inhibition placed upon this language.

Two contributions from our collection focus on investigating *reading processes* using *eye-tracking*. Lou et al. suggest that eye

movements during reading can be influenced by the *motivation of self-enhancement* in addition to various stimulus' properties and cognitive factors; this also indicates that eye-tracking can be used to study implicit social cognition. Research presented by Petrova et al. shows that *readers process information better and faster while reading sketch-notes than verbal texts*; additionally, various types of sketch-notes differ in terms of how good the readers are in following the order of elements.

Finally, two articles offer examples of *behavioral psycholinguistic* research. Niebuhr et al. report the results of a 12-weeks *prosodic charisma training* that is shown to be more beneficial for female speakers as opposed to male ones. Pokhoday et al. report new evidence about the role of *the speaker's attention* (manipulated by visual priming) and event orientation in sentence production by using a flexible word-order language, Russian.

In conclusion, the present Research Topic will undoubtedly contribute to a better understanding of how neurocognitive systems provide humans with language and will help to further unveil the backstage of our intrinsic communication abilities.

AUTHOR CONTRIBUTIONS

OS and YS conceived the paper and the Research Topic. OS, YS, AM, and BM-L contributed to the final version of the manuscript. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

This issue became possible as a result of a series of international meetings on Neurobiology of Speech and Language organized by the Laboratory of Behavioural Neurodynamics of Saint Petersburg State University. This work was financially supported by the Government of Russian Federation (grant contract no. 14.W03.31.0010).

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Shcherbakova, Myachykov, Martín-Luengo and Shtyrov. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Effects of Self-Enhancement on Eye Movements During Reading

Ya Lou^{1,2,3}, Huajian Cai^{1,2}, Xuwei Liu^{1,2} and Xingshan Li^{1,2*}

¹ CAS Key Laboratory of Behavioral Science, Institute of Psychology, Beijing, China, ² Department of Psychology, University of Chinese Academy of Sciences, Beijing, China, ³ Beijing Institute of Education, Beijing, China

Previous studies show that readers' eye movements are influenced by text properties and readers' personal cognitive characteristics. In the current study, we further show that readers' eye movements are influenced by a social motivation of self-enhancement. We asked participants to silently read sentences that describe self or others with positive or negative traits while their eyes were monitored. First-fixation duration and gaze duration were longer when positive words were used to describe self than to describe others, but there was no such effect for negative words. These results suggest that eye movements can be influenced by the motivation of self-enhancement in addition to various stimuli features and cognitive factors. This finding indicates that the eye movement methodology can potentially be used to study implicit social cognition.

Keywords: self enhancement, reading, eye movements, language sciences, personality and social psychology

OPEN ACCESS

Edited by:

Andriy Myachykov,
Northumbria University,
United Kingdom

Reviewed by:

Stefan Hawelka,
University of Salzburg, Austria
Kevin B. Paterson,
University of Leicester,
United Kingdom

*Correspondence:

Xingshan Li
lixs@psych.ac.cn

Specialty section:

This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Psychology

Received: 01 November 2018

Accepted: 04 February 2019

Published: 25 February 2019

Citation:

Lou Y, Cai H, Liu X and Li X (2019)
Effects of Self-Enhancement on Eye
Movements During Reading.
Front. Psychol. 10:343.
doi: 10.3389/fpsyg.2019.00343

INTRODUCTION

We move our eyes three to four times per second when we are awake to selectively perceive visual information that is most salient or most relevant to our current task (Rayner, 1998, 2009). Decades of eye movement research has shown that our eye movements are influenced by various features of visual stimuli (e.g., words frequency in reading) (Rayner, 1998, 2009) and diverse personal cognitive characteristics (e.g., reading skills) (Rayner and Reichle, 2010; Kuperman and Dyke, 2011). For example, high-frequency words are typically fixated for less time and skipped more often than low-frequency words (Inhoff and Rayner, 1986; Rayner and Duffy, 1986); skilled readers make shorter fixations on words, longer forward saccades (Jared et al., 1999; Ashby et al., 2005), and fewer regressions compared to less-skilled readers (Ashby et al., 2005). Moreover, the eye movements also vary with purposes. For example, Kaakinen et al. (2002) examined the effects of reading goals on eye movement behavior. The reading goal was induced by instructing the participants to imagine that they were going to live in another country. Then the participants were asked to read an expository text that included four remote countries. They found that readers made more and longer fixations on sentences that described the conditions of that country than on other sentences. The current research aims to explore whether eye movement in reading could be influenced by a special human motivation, that is, the desire for a positive self or self-enhancement.

Word processing time might also be affected by high-level cognition such as motivation. For example, self-enhancement, a type of motivation that works to make people feel good about themselves, might affect eye movements during reading. Self-enhancement makes people favor positive over negative self-views (Sedikides and Gregg, 2008). A positive self is significant for

physical and mental health (Taylor and Brown, 1988), well-being (Baumeister et al., 2003), and coping with threats and life difficulties (see Alicke and Sedikides, 2009 for a review). Self-positivity may manifest on various behavioral indexes, such as trait endorsement (D'Argembeau et al., 2005; Kwan et al., 2007), reaction time (Paulhus and Levitt, 1987; Gebauer et al., 2012), and neural responses, such as electroencephalograph (EEG) signal (Luck, 2005; Cai et al., 2016; Hampton and Varnum, 2017). Therefore, self-enhancement might influence people's attention allocation during reading, favoring positive information about self.

This research explored whether self-enhancement manifests in eye movements in reading, or whether self-enhancement influences eye movements in reading self-relevant information. Accordingly, the participants were asked to silently read sentences that describe the self or others with positive or negative traits while their eyes were monitored. Each sentence contained one identity word (i.e., I or He) and one attribute word (i.e., positive or negative) (see **Figure 1** for examples). Previous studies showed that people tend to judge positive personality attributes to be more appropriate in describing themselves than in describing others and therefore self-enhancement may encourage people to "elaborate, dwell on" positive self-evaluative information (Heine et al., 1999, p. 760). We inferred that positive traits that describe the self may obtain longer fixation time than those describing the other person (i.e., *he*).

MATERIALS AND METHODS

Participants

A total of 40 undergraduate students from Beijing Forestry University and China Agricultural University participated in this experiment. Three participants were excluded because of technical problems or track loss during eye movement recording. Participants provided consent in accordance with the protocols approved by the ethics committee of Institute of Psychology, Chinese Academy of Sciences.

Apparatus

Eye movements were recorded using an SR-Research Eyelink 1000 eye tracker (Kanata, ON, Canada) sampling at a rate of 1,000 Hz. Eye movements were recorded from the right eye during binocular viewing. The sentences were displayed as a single line of text using 24 point Song font. The participants were seated at a distance of 58 cm from the computer monitor.

Materials and Design

The trait words were adapted from a previous study (Cai, 2003), and comprised 12 positive words and 12 negative words (see **Table 1**). The average frequency of the positive trait words ($M = 45.53$ occurrences per million words, $SD = 56.69$) was higher than that of the negative ones ($M = 8.59$ occurrences per million words, $SD = 10.95$). Each trait word was embedded in two different sentence frames with the following subjects in the sentences: one with the embedded word "I" preceding the trait word and the other with "He." The word "I" and "He" were

used as identity words. Therefore, 12 sentences were created for each of the following four conditions: *I-positive*, *He-positive*, *I-negative*, and *He-negative* (see **Figure 1**). The average sentence length ranged from 16 to 30 characters with a mean of 20.85 characters and a standard deviation of 3.45. The same number of sentences was created as filler sentences in which neither identity nor trait word was included.

Procedure

The participants were tested individually. When they arrived at the lab, they were informed that this experiment was designed to use an eye tracking technology to investigate sentence-comprehension processes. However, they were unaware of the experiment's purpose. Thereafter, they performed a calibration procedure by looking at a sequence of three fixation points that were randomly displayed horizontally across the middle of the computer screen. The maximal calibration error was 0.5° . Calibration was conducted at the beginning of the experiment and was conducted again during the experiment when necessary. At the beginning of each trial, a drift check was conducted to ensure that the error of the eye tracker was within the allowable range. Thereafter, the participants looked at a square located at the position of the first character of the sentence. After they fixated at this square for 0.5 s, the entire sentences appeared. The participants silently read the sentences, and they were required to press a button when they had completed reading these sentences. A comprehension question with a two-alternative forced-choice response was asked after each of all the 24 filler items and participants responded by pressing one of two keys on a response box. These questions were created to ensure that the participants carefully read the sentences. The mean accuracy of the comprehension questions was 95%, thereby indicating that the participants carefully read the sentences.

Data Analysis

Fixations above 1000 ms or below 80 ms were excluded from analyses. We report the following eye movement measures for the target words in the sentences (Rayner, 1998): (a) First-fixation duration (duration of the first first-pass fixation on the target word), (b) Gaze duration (sum of all first-pass fixations on the target word prior to proceeding to another word), (c) Skipping probability (the probability that the target word was skipped on first-pass reading), and (d) Total reading time (sum of all fixations on the target word, including regression). First-fixation duration and gaze duration are sensitive to early processing associated with lexical identification, whereas total reading times are sensitive to later processes associated with integration (Inhoff, 1984). **Table 2** presents the descriptive statistics of these eye movement measures.

Given that high-frequency words are processed faster than low-frequency words (known as frequency effect) (Rayner and Duffy, 1986), and the frequency of the positive trait words were higher than those of the negative trait words, and the comparison between negative and positive words was not relevant to our research question, we did not directly compare eye movement measures between the negative words and positive words.

Materials used in the experiment.

I-positive Condition

我幼时乖巧伶俐，从三岁起就随父亲练字。

I had been a **clever** child; I studied calligraphy with my farther starting at three-years old.

He-positive Condition

他口齿伶俐，每次发言都滔滔不绝。

He is an **articulate** man and not afraid to speak his mind.

I-negative Condition

我在跳舞时动作笨拙，蹒跚摇晃，险些儿跌跤。

I felt **awkward** when I was dancing, I staggered and swayed and nearly fell down.

He-negative Condition

他口才笨拙，不流利的发言一度成为全校笑柄。

His **unprofitable** eloquence at schoolwide banquets has been the butt of many jokes

FIGURE 1 | Materials used in experiment. The identify words are underlined and the trait words are in bold letters for the purpose of illustration (the characters were neither underlined nor made bold in the experiment).

TABLE 1 | Trait words used in this experiment.

Trait words						
Positive	聪明	漂亮	成功	价值	高尚	强壮
	(smart)	(pretty)	(successful)	(worthy)	(noble)	(robust)
Negative	自豪	可爱	诚实	受尊重	有能力	伶俐
	(proud)	(lovely)	(honest)	(respectful)	(capable)	(clever)
Negative	愚蠢	丑陋	失败	讨厌	无能	卑鄙
	(stupid)	(ugly)	(unsuccessful)	(disgusting)	(incapable)	(contemptible)
Negative	罪恶	笨拙	堕落	可恨	虚弱	可耻
	(evil)	(stupid)	(corrupt)	(hateful)	(feebleness)	(disgraceful)

Instead, the key comparisons were the results between the *I-positive* and *He-positive* conditions and between the *I-negative* and *He-negative* conditions.

Eye movement data were analyzed using linear mixed-effects models (LMM) for continuous variables (Baayen et al., 2008; Jaeger, 2008), in which the participants and items were considered as random effects. Identity words, trait words, and their interactions were entered as fixated effects. The analyses were performed using the *lme4* package (Bates et al., 2014) in the R statistical software (Version 3.3.1, R Core Team, 2016), and the *lmerTest* Package was used to get the *p*-value for tests for fixed effects.

RESULTS

First-Fixation Duration

First-fixation durations were shorter in the positive condition ($M = 223$ ms, $SE = 4$) than in the negative condition

($M = 229$ ms, $SE = 4$), $b = -22.575$, $SE = 10.279$, $t = -2.196$, $p = 0.03$. No difference was observed between the identity conditions (the *I* condition: $M = 228$ ms, $SE = 4$, the *He* condition: $M = 224$ ms, $SE = 4$), $t = -1.410$, $p = 0.158$. However, the interaction effect between the trait

TABLE 2 | Eye movement measures in the trait word region.

	Positive				Negative			
	I		HE		I		HE	
First-fixation duration	231	(5)	214	(5)	226	(5)	233	(6)
Skipping probability	0.28	(0.02)	0.33	(0.02)	0.23	(0.02)	0.25	(0.02)
Gaze duration	284	(12)	257	(10)	259	(8)	261	(8)
Total reading time	368	(15)	373	(15)	332	(12)	334	(12)
Regression in	0.12	(0.02)	0.12	(0.02)	0.10	(0.01)	0.10	(0.01)

First-fixation duration, gaze duration, and total time were measured in ms. SEs are provided in parentheses.

valence and identity was significant, $b = 28.591$, $SE = 10.516$, $t = 2.719$, $p < 0.01$.

Planned comparisons showed that first-fixation durations in the *He-positive* condition ($M = 214$ ms, $SE = 5$) were shorter than those in the *I-positive* condition ($M = 231$ ms, $SE = 5$), $b = 17.795$, $SE = 7.136$, $t = 2.494$, $p = 0.01$. First-fixation duration on the negative trait words did not differ between the *He-negative* condition ($M = 233$ ms, $SE = 6$) and *I-negative* condition ($M = 231$ ms, $SE = 5$), $t = -1.2$, $p = 0.2$.

Gaze Duration

The interaction effect of the trait words and identity was not significant, $b = 34.407$, $SE = 18.207$, $t = 1.890$, $p = 0.06$. No main effect of the trait word and identity was observed, both $t < 1$. Since the interaction was close to significant, we also conducted some further exploratory analyses. Planned comparisons showed that gaze durations were shorter in the *He-positive* ($M = 258$ ms, $SE = 10$) than those in the *I-positive* ($M = 284$ ms, $SE = 12$) condition, $b = 27.56$, $SE = 14.31$, $t = 1.93$, $p = 0.05$. Gaze duration on the trait words did not differ between the *He-negative* ($M = 261$ ms, $SE = 8$) and *I-negative* condition ($M = 259$ ms, $SE = 8$), $t < 1$. The pattern of gaze duration replicated the results from the first-fixation duration.

Other Measures

Neither the main effects of the trait words, identity, nor the interaction were significant for the other measures (skipping probability, total time, all $t < 1$).

DISCUSSION

This study analyzed whether human motivation, particularly self-enhancement, influences eye movements during reading. The identity words (i.e., I versus He) and trait words (i.e., positive versus negative) were embedded in the sentences. Accordingly, four sentence conditions (i.e., *I-positive*, *He-positive*, *I-negative*, and *He-negative*) were created. As expected, we found that first-fixation duration and gaze duration in the *I-positive* condition were longer than those in the *He-positive* condition. However, we found no difference in fixation time on negative words.

These findings showed that self-enhancement can affect eye movement behavior during reading. To enhance or maintain a positive self, people often selectively remember their strengths rather than weaknesses. One way to do this is at the encoding stage of memory through selective attention. As a result, people dwell longer on positive words that describe self in reading. These results suggest that eye movements are affected by reading-related factors (e.g., reading material features and reading ability) and human motivation (e.g., self-enhancement), thereby extending our understanding of the range of factors that can affect mechanisms of eye movement control during reading.

For negative traits, we did not observe shorter fixation time on negative words that describing *I* than those that describe *he*. Two factors might have jointly affected the processing of negative words. First, negative words that describing *I* might be processed for shorter time than those that describe *he* due to the need

for self-protection. However, there may be considerably longer fixation time because negative self-information may constitute a conflict with the existing self-positivity, thereby attracting further attention due to its inconsistency or novelty. These two opposite factors might have caused a small difference (or no difference) in fixation time between the two *negative* conditions.

In addition to enhancing our understanding of eye movement control during reading, our findings also suggest that eye movement methodology can be used to study the on-line effects of self-positivity during comprehension. Previous studies have used self-report scales (Rosenberg, 1965), reaction time task (e.g., Implicit Association Test or IAT, Greenwald et al., 1998), and electroencephalograph EEG signal (Luck, 2005; Wu et al., 2016) to measure self-positivity. Compared with other methodologies, eye tracking technique has a few advantages. First, this technique reflects moment-to-moment cognitive processes without interfering with the natural behavior of the participants. Second, eye movement data provide the researcher with valuable temporal information about exactly when a manipulation exerts influences. Moreover, the task is based on spontaneous reactions, thereby possibly assisting in sidestepping many artifacts, such as social desirability and response styles.

There are some limitations in the current study. First, we only used limited number of stimuli. Further studies are needed to investigate whether the effects observed in the current study can be extended to other types of positive and negative words. Second, we did not directly compare the effects that were observed with eye tracking technology with the findings that were observed using other technologies. Further studies are needed to address these issues.

In summary, we showed that eye movements can be influenced by the motivation of self-enhancement beyond various stimuli features and cognitive factors. This finding broadens our understanding of the sensitivity of eye movements to high level cognitive processes by showing that differences in the processing of self versus other descriptive words are detectable in early processing using eye movement measures.

AUTHOR CONTRIBUTIONS

XiL and YL designed the experiments. YL and XuL carried out experiments and analyzed the experimental results. YL, XiL, XuL, and HC wrote the manuscript.

FUNDING

This research was jointly funded by the National Natural Science Foundation of China (NSFC) and the German Research Foundation (DFG) in Project Crossmodal Learning, NSFC 61621136008/DFG TRR-169. This research was supported by a grant from the National Natural Science Foundation of China (31571125) and a grant from the National Social Science Fund of China (17ZDA305). This research was also supported by CAS Key Laboratory of Behavioral Science, Institute of Psychology.

REFERENCES

- Alicke, M. D., and Sedikides, C. (2009). Self-enhancement and self-protection: what they are and what they do. *Eur. Rev. Soc. Psychol.* 20, 1–48. doi: 10.1080/10463280802613866
- Ashby, J., Rayner, K., and Clifton, C. (2005). Eye movements of highly skilled and average readers: differential effects of frequency and predictability. *Q. J. Exp. Psychol. A* 58, 1065–1086. doi: 10.1080/02724980443000476
- Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *J. Mem. Lang.* 59, 390–412. doi: 10.1016/j.jml.2007.12.005
- Bates, D., Maechler, M., Bolker, B., and Walker, S. (2014). *lme4: Linear Mixed-Effects Models Using Eigen and S4. R Package Version*. Available at: https://www.researchgate.net/publication/279236477_Package_Lme4_Linear_Mixed-Effects_Models_Using_Eigen_and_S4
- Baumeister, R. F., Campbell, J. D., Krueger, J. I., and Vohs, K. D. (2003). Does high self-esteem cause better performance, interpersonal success, happiness, or healthier lifestyles? *Psychol. Sci. Public Interest* 4, 1–44. doi: 10.1111/1529-1006.01431
- Cai, H. (2003). Nei yin zi zun ji nei yin zi zun yu wai xian zi zun de guan xi [The implicit self-esteem among Chinese in relation to explicit self-esteem]. *Acta Psychol. Sin.* 35, 796–801.
- Cai, H., Wu, L., Shi, Y., Gu, R., and Sedikides, C. (2016). Self-enhancement among Westerners and Easterners: a cultural neuroscience approach. *Soc. Cogn. Affect. Neurosci.* 11, 1569–1578. doi: 10.1093/scan/nsw072
- D'Argebeau, A., Comblain, C., and Van der Linden, M. (2005). Affective valence and the self-reference effect: influence of retrieval conditions. *Br. J. Psychol.* 96, 457–466. doi: 10.1348/000712605X53218
- Gebauer, J. E., Görtz, A. S., Hofmann, W., and Sedikides, C. (2012). Self-love or other-love? Explicit other-preference but implicit self-preference. *PLoS One* 7:e41789. doi: 10.1371/journal.pone.0041789
- Greenwald, A. G., McGhee, D. E., and Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *J. Personal. Soc. Psychol.* 74:1464. doi: 10.1037/0022-3514.74.6.1464
- Hampton, R. S., and Varnum, M. (2017). Do cultures vary in self-enhancement? ERP, behavioral, and self-report evidence. *Soc. Neurosci.* 13, 566–578. doi: 10.1080/17470919.2017.136147
- Heine, S. J., Lehman, D. R., Markus, H. R., and Kitayama, S. (1999). Is there a universal need for positive self-regard? *Psychol. Rev.* 106, 766–794.
- Inhoff, A. W. (1984). Two stages of word processing during eye fixations in the reading of prose. *J. Verbal Learning Verbal Behav.* 23, 612–624. doi: 10.1016/S0022-5371(84)90382-7
- Inhoff, A. W., and Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: effects of word frequency. *Percept. Psychophys.* 40, 431–439. doi: 10.3758/BF03208203
- Jaeger, T. F. (2008). Categorical data analysis: away from ANOVAs (transformation or not) and towards logit mixed models. *J. Mem. Lang.* 59, 434–446. doi: 10.1016/j.jml.2007.11.007
- Jared, D., Levy, B. A., and Rayner, K. (1999). The role of phonology in the activation of word meanings during reading: evidence from proofreading and eye movements. *J. Exp. Psychol. Gen.* 128, 219–264. doi: 10.1037/0096-3445.128.3.219
- Kaakinen, J. K., Hyönä, J., and Keenan, J. M. (2002). Perspective effects on online text processing. *Dis. Process.* 33, 159–173. doi: 10.1207/S15326950DP3302_03
- Kuperman, V., and Dyke, J. A. V. (2011). Effects of individual differences in verbal skills on eye-movement patterns during sentence reading. *J. Mem. Lang.* 65, 42–73. doi: 10.1016/j.jml.2011.03.002
- Kwan, V. S., Barrios, V., Ganis, G., Gorman, J., Lange, C., Kumar, M., et al. (2007). Assessing the neural correlates of self-enhancement bias: a transcranial magnetic stimulation study. *Exp. Brain Res.* 182, 379–385. doi: 10.1007/s00221-007-0992-2
- Luck, S. J. (2005). *An Introduction to the Event-Related Potential Technique*. Cambridge, MA: MIT Press.
- Paulhus, D. L., and Levitt, K. (1987). Desirable responding triggered by affect: automatic egotism? *J. Personal. Soc. Psychol.* 52, 245–259. doi: 10.1037/0022-3514.52.2.245
- R Core Team (2016). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Available at: <https://www.R-project.org/>
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychol. Bull.* 124, 372–422. doi: 10.1037/0033-2909.124.3.372
- Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Q. J. Exp. Psychol.* 62, 1457–1506. doi: 10.1080/17470210902816461
- Rayner, K., and Duffy, S. A. (1986). Lexical complexity and fixation times in reading: effects of word frequency, verb complexity, and lexical ambiguity. *Mem. Cogn.* 14, 191–201. doi: 10.3758/BF03197692
- Rayner, K., and Reichle, E. D. (2010). Models of the reading process. wiley interdisciplinary reviews. *Cogn. Sci.* 1, 787–799.
- Rosenberg, M. (1965). *Society and the Adolescent Self-Image*. Princeton, NJ: Princeton university press. doi: 10.1515/9781400876136
- Sedikides, C., and Gregg, A. P. (2008). Self-enhancement: food for thought. *Perspect. Psychol. Sci.* 3, 102–116. doi: 10.1111/j.1745-6916.2008.00068.x
- Taylor, S. E., and Brown, J. D. (1988). Illusion and well-being: a social psychological perspective on mental health. *Psychol. Bull.* 103, 193–210. doi: 10.1037/0033-2909.103.2.193
- Wu, L., Gu, R., Cai, H., and Zhang, J. (2016). Electrophysiological evidence for executive control and efficient categorization involved in implicit self-evaluation. *Soc. Neurosci.* 11, 153–163. doi: 10.1080/17470919.2015.1044673

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Lou, Cai, Liu and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Embodied Semantics in a Second Language: Critical Review and Clinical Implications

Elisa Monaco^{1*}, Lea B. Jost¹, Pascal M. Gyga² and Jean-Marie Annoni^{1,3}

¹ Laboratory for Cognitive and Neurological Sciences, Neurology Unit, Medicine Section, Department of Neuroscience and Movement Science, Faculty of Science and Medicine, University of Fribourg, Fribourg, Switzerland, ² Department of Psychology, University of Fribourg, Fribourg, Switzerland, ³ Neurology Unit, Fribourg Cantonal Hospital, Fribourg, Switzerland

The role of the sensorimotor system in second language (L2) semantic processing as well as its clinical implications for bilingual patients has hitherto been neglected. We offer an overview of the issues at stake in this under-investigated field, presenting the theoretical and clinical relevance of studying L2 embodiment and reviewing the few studies on this topic. We highlight that (a) the sensorimotor network is involved in L2 processing, and that (b) in most studies, L2 is differently embodied than L1, reflected in a lower degree or in a different pattern of L2 embodiment. Importantly, we outline critical issues to be addressed in order to guide future research. We also delineate the subsequent steps needed to confirm or dismiss the value of language therapeutic approaches based on embodiment theories as a complement of speech and language therapies in adult bilinguals.

Keywords: aphasia, bilingualism, clinical rehabilitation, embodiment, semantics

OPEN ACCESS

Edited by:

Yury Y. Shtyrov,
Aarhus University, Denmark

Reviewed by:

Adolfo M. García,
Laboratory of Experimental
Psychology and Neuroscience,
Argentina
Véronique Boulenger,
UMR5596 Dynamique du Langage,
France

*Correspondence:

Elisa Monaco
elisa.monaco@unifr.ch

Received: 05 November 2018

Accepted: 12 March 2019

Published: 29 March 2019

Citation:

Monaco E, Jost LB, Gyga PM
and Annoni J-M (2019) Embodied
Semantics in a Second Language:
Critical Review and Clinical
Implications.
Front. Hum. Neurosci. 13:110.
doi: 10.3389/fnhum.2019.00110

INTRODUCTION

The term “embodiment” refers to the grounding of cognition in systems involved in low level perceptual and action information processing. Embodied theories of cognition claim that higher cognitive processing, including language, activates the same brain sensorimotor structures involved when experiencing the environment (e.g., Glenberg, 1997; Glenberg and Kaschak, 2002; Gallese and Lakoff, 2005; Pulvermüller et al., 2005; Barsalou, 2008; Jirak et al., 2010; Meteyard et al., 2012). Converging clinical and neurophysiological evidence indicates that semantic knowledge is grounded in different heteromodal but also on modality specific cortical regions, coding for perceptual, sensory, visual, auditory, motor or affective experiential information. This distributed network coding for conceptual processes has also been called “experiential brain system” (e.g., Ghio and Tettamanti, 2016).

The idea that language processing activates sensorimotor areas of the brain has been supported by neuroimaging and neuromodulation studies focusing on the processing of nouns, adjectives, verbs and sentences including actions performed by specific body parts or manipulable objects. These studies suggested that primary and secondary motor cortices were regularly involved (Hauk et al., 2004; Buccino et al., 2005; Pulvermüller et al., 2005; Tettamanti et al., 2005; Aziz-Zadeh et al., 2006; Boulenger et al., 2009; Papeo et al., 2009; Alemanno et al., 2012; Gough et al., 2013; Innocenti et al., 2014; Gianelli and Dalla Volta, 2015). Similarly, in studies on emotion, mimetic muscles have been shown to react to emotional words and sentences (Havas et al., 2007;

Foroni and Semin, 2009, 2013; Havas et al., 2010; Davis et al., 2015; Foroni, 2015; Fino et al., 2016; Baumeister et al., 2017). Others have also shown correlations between the impairment in action word processing (e.g., “to pour,” “to wave”) and the impairment in action performance, assessed using a visually guided reaching task (e.g., Desai et al., 2015). Finally, some have also shown that a virtual transient lesion induced by repetitive Transcranial Magnetic Stimulation (TMS) or transcranial Direct Current Stimulation (tDCS) over the premotor and motor cortex affects comprehension of action related language (e.g., Willems et al., 2011; Tremblay et al., 2012; Vukovic et al., 2017; Gijssels et al., 2018).

In the clinical setting – following the seminal work by Warrington and McCarthy (1987) – the idea that different “weighting values” from independent perceptual channels could subserve different categories of knowledge is rather undisputed. In their paper, Warrington and McCarthy (1987) presented a severe dysphasic patient who showed impairment in selecting objects (and not food or animate beings) as well as specifically small manipulable objects (and not large man-made objects). Later clinical studies confirmed the interaction between language processing and the activation of perceptuo- and sensori-motor brain areas. For example, Arévalo et al. (2007) showed that manipulable words (e.g., “comb,” “kite”) were distinct to non-manipulable ones (e.g., “smoke,” “moon”), not only behaviorally, but also in their associated activated brain areas. Others have shown that lesions to the sensorimotor areas were associated with impaired processing of lexical and conceptual knowledge of actions (e.g., Kemmerer et al., 2012). In fact, sensorimotor network impairment – due to neurodegenerative diseases – has been shown to selectively compromise the processing of action verbs, motor-language coupling, syntax, and the processing of graspable objects (e.g., Bak et al., 2006; Cotelli et al., 2007; Cardona et al., 2013; Fernandino et al., 2013a,b; Kargieman et al., 2014; Birba et al., 2017; Buccino et al., 2017a; Cotelli et al., 2018). Note that (1) these effects seem to be independent of the general cognitive functioning and of the actual manifestation of the symptoms (e.g., Bocanegra et al., 2015, 2017; García et al., 2017), and that (2) they have not always been found. For example, in some studies, lesions to the motor cortex did not cause deficits in action word processing (e.g., Papeo et al., 2010; Maieron et al., 2013)¹. Studies such as these do question the very necessity of activating sensorimotor structures when processing language. They reflect the idea that although embodied cognition is an interesting concept, it is unlikely that *all* our cognition is grounded in sensorimotor experiences (Goldinger et al., 2016). In fact, most contemporary embodied theories do claim that grounded cognition *complements* existing accounts, without the presumption of replacing them, yet it offers new opportunities to study basic cognitive processes (Barsalou, 2016). Hence, despite conceptual controversies (e.g., Mahon and Caramazza, 2008; Papeo et al., 2013; Caramazza et al., 2014; Martin, 2016), the idea that perceptuo- and sensori-motor information is activated when semantic representations

are accessed (Meteyard et al., 2012) is extremely interesting in terms of bilingualism and clinical implications. In terms of the former implications, a central issue has been whether lexico-semantic representations are shared or distinct between L2 and L1 (the mother tongue). In terms of the latter, if L2 is less (or not) embodied, clinicians – often confronted to patients whose first language is not the language of rehabilitation – could choose different therapy strategies (more related to action observation or gestures) in L1 but not in L2. We strongly believe that understanding how both languages are represented in the brain and how they interact with one another will help diagnosing and optimizing rehabilitation strategies and health care. To our knowledge, only two other reviews have discussed embodiment and bilingualism: one focusing on emotion studies (Pavlenko, 2012), and one theoretical paper discussing embodiment predictions in bilingualism and presenting clinical implications for children with a Developmental Language Disorder (Adams, 2016). In the present review we wish to further the latter and stress the relevance of studying embodiment in L2 by (1) discussing bilingual language models from this perspective, (2) presenting studies that have linked L2 and embodiment, and (3) calling attention to the concrete clinical implications of the processes at stake.

Note that to keep the focus of the present paper specifically on embodiment and second language lexico-semantic representations and processing (and the subsequent clinical implications), we only briefly mention the work on embodiment *while acquiring* a second language. Although slightly satellite to the present concerns, research on the latter has also raised some important issues for bilingualism research (see for example Macedonia, 2014; Wellsby and Pexman, 2014; Buccino and Mezzadri, 2015; Macedonia and Mueller, 2016)².

BILINGUAL LANGUAGE MODELS AND EMBODIMENT

The Influence of Proficiency, Immersion and Age of Acquisition on Semantic Representations in Bilingual Models

Current models of bilingualism assume that, when processing a word (either in L1 or L2), after an initial language specific visual processing (Khateb et al., 2016), associated lexico-semantic is activated for both languages (e.g., van Heuven and Dijkstra, 2010; Moon and Jiang, 2012). They also assume that the parallel activation of the two languages is modulated by subject-related factors, such as age of acquisition (AoA; i.e., the age at which bilinguals begin to learn L2, Hernandez and Li, 2007), L2 exposure and/or L2 proficiency. Importantly though, these models do differ in the way they conceptualize lexico-semantic systems. The Revised Hierarchical model (RHM) (Kroll and Stewart, 1994), for example, assumes that each language has a specific lexical system, yet both languages share semantic representations that are stored in a common memory system.

¹Taylor et al. (2017) argued that in those studies, action and motion were not considered separately, leading to erroneous interpretations.

²Other examples of studies on L2 acquisition and embodiment are provided in the Section “Critical Synthesis.”

In this framework, L2 to L1 connections are more developed than vice-versa, but with increasing proficiency, the strength of L2 connections changes. Other models, such as the Bilingual Interactive Activation Plus model (BIA+) by Dijkstra and van Heuven (2002), assume that the lexical representations of the two languages are somehow integrated. As such, access to the orthographic, phonologic, or semantic representations is non-selective between languages. Dijkstra and van Heuven (2002) and van Heuven and Dijkstra (2010) further discuss how the proficiency of a language relies on the frequency of word usage. As such, it is linked to the rapidity by which those words' representations are activated. Therefore, in case of low L2 proficiency, the authors argue for a temporal delay to access representations in L2 compared to L1.

In addition to language proficiency, more global exposure to L2 environment plays a role in semantic processing (e.g., Perani et al., 2003; Stein et al., 2014), although these factors are most likely interdependent. Exposure increases proficiency, even to the extent – in extreme cases – of hindering lexical access in L1 (e.g., Linck et al., 2009). Similarly, L2 proficiency is linked to age of acquisition (e.g., Johnson and Newport, 1989). However, L2 proficiency and AoA have been suggested to have different roles in language processing. In particular, language proficiency seems to be more influential than AoA in semantic processes, while AoA would rather play a role in syntactic knowledge (Wartenburger et al., 2003; Abutalebi, 2008). Some have questioned this assumption (e.g., Izura and Ellis, 2004; Isel et al., 2010; Sabourin et al., 2014), suggesting that AoA's influence was also on the lexico-semantic level. This is in line with the model advanced by Silverberg and Samuel (2004), which postulates a common semantic system between languages only in the case of early AoA. The conceptual environment may be similar, yet only if the two languages are acquired at a similar age. For late L2 learners, the conceptual context has been shaped by years of experiences in L1. This vision is similar in the Sense Model (Finkbeiner et al., 2004), which postulates that L2 lexical semantic representations have less “senses” associated with them in comparison to those in L1.

Semantic representations in bilingual models are therefore differently influenced by proficiency, exposure and age of acquisition, all factors to be taken into account in predicting embodiment in L2.

Embodiment Predictions for L2 and Their Impact on Language Models

Despite some evidence suggesting sensorimotor involvement in L1 semantic processing, to our knowledge, only few studies have investigated such involvement in L2 processing. The lack of studies on the topic could be explained by two different, yet related assumptions. First, when considering early bilinguals, given that both languages are learnt in the same cultural context³, one could assume an overlap of sensorimotor information between the two different languages (Adams, 2016). Second, and contrariwise, in late bilinguals, L2 is often acquired explicitly

in a school context, hence without a true involvement of sensory modalities. As such, sensorimotor activation in the two languages should be different, with less rich or direct connections to the sensorimotor cortex for the second language (Perani and Abutalebi, 2005; Pavlenko, 2007; Eilola and Havelka, 2011; Dudschig et al., 2014; see also Declarative/Procedural model on implicit and explicit language learning, Aglioti, 1999; Ullman, 2001, 2004; Paradis, 2004; Hamrick et al., 2018). Yet, if semantic representations are shared between L1 and L2, as assumed by some models of bilingualism, we should not expect a difference in the embodiment of the two languages. One could also argue that in moderately proficient bilinguals (and late AoA), the link between the L2 lexical store and the semantic system is most likely not as developed as that of L1. Consequently, such a weaker connection could translate to different embodiment effects in L2.

Transferring this assumption into clinical predictions, the assessment and rehabilitation of a patient in L2 – acquired late and/or less proficient – could depend on the patient's embodiment of L1 as well as the possible transfer between languages. It could also depend on the way the two languages are stored. Even if – as assumed by models considering separate stores of concepts for both languages (e.g., Finkbeiner et al., 2004; Silverberg and Samuel, 2004) – the path to access semantic representations is not influenced by a delayed access through L1, the strength of connections between semantics and sensorimotor structures could still vary. Consequently, from a clinical standpoint, both the assessment and the therapy of the lexico-semantic system could be different depending on the language at hand (i.e., L1 or L2). Namely, although specific language tasks may constitute potential markers for movement disorders in L2 – as they do in L1 (e.g., Cardona et al., 2013; Birba et al., 2017; García et al., 2017, 2018) –, this would only be the case if L2 was grounded in the motor system (see section Motor-Language Interactions) and it may depend on its actual degree of embodiment. In the same vein, any transfer of therapy improvement from one language to another is more likely if the same linguistic processes are targeted, such as lexical or phonological encoding (e.g., Laganaro and Overton Venet, 2001). The transfer of outcomes from L1 to L2 would hence be larger if semantic representations are shared, as suggested by some of the bilingual models discussed earlier (e.g., Dijkstra and van Heuven, 2002).

Investigating the sensorimotor activation in L2 – and its therapeutic context – could also offer some insight on models of L1, providing further understanding of the timing of sensorimotor involvement in language processing. Besides this debate (e.g., Mahon and Caramazza, 2008; Postle et al., 2013), answering such a question could generally help us to understand the role of sensorimotor language therapies. We could even argue that a better grasp of the involvement of sensorimotor structures in both L1 and L2 could further models of language representation as well as models of motor-language coupling (e.g., HANDLE, García and Ibáñez, 2016) and of language acquisition (e.g., ABL model of Glenberg and Gallese, 2012). In fact, some language acquisition and development models have already taken embodiment evidence into account. For example, the Word as Social Tool (WAT) model

³Here we do not discuss the case of bicultural bilinguals, but readers can refer to Jared et al. (2013) and Adams (2016) for a discussion on this subject.

(Borghi and Cimatti, 2009) considers words not only as a referent, but also as a tool to operate in the world. This model already posits different modes of acquisition, namely perceptual, linguistic or mixed, with the level of embodiment depending on these modes (Scorilli et al., 2011). Such a model could be helpful in making predictions for future research in L2 learning and recovery. As an example, while acquiring a language, or in language therapies fostering interactions (e.g., CIAT by Pulvermüller et al., 2001), L2 could be better acquired or retrieved with increasing amounts of social or embodied experiences.

STUDIES ON L2 EMBODIED SEMANTIC IN HEALTHY POPULATIONS

Although we have, so far, only presented L2 and embodiment as predictions and conjectures, some studies specifically addressing this issue do exist, and we critically discuss them next, raising some of the remaining open questions not yet answered. To facilitate a global perspective on those studies, we present in **Tables 1, 2** summaries of their methodological, theoretical and interpretative essence.

Behavioral Studies

Bergen et al. (2010) assessed sensorimotor activation when native (Experiment 1) and non-native English speakers (Experiment 4) process words in English. In their task, participants had to indicate if a written verb was or was not a good description for an action depicted in a preceding image. The verbs could either match the image (e.g., an image of someone running with the verb *run*) or mismatch it, yet refer to actions using the same (e.g., *kick*) or a different effector (e.g., *drink*). In the mismatch condition, participants were slower to correctly respond when the verb used the same effector than when it was different. This interference effect was similar for non-native and native English speakers, suggesting that both groups relied on sensorimotor activation to understand verbs. Still, English proficiency, calculated as accuracy in the task, was positively correlated with the size of the effect (Bergen et al., 2010).

In a similar vein, in Buccino et al. (2017b), Italian students performed a *go-no go* task in which English nouns and pictures of graspable and non-graspable objects were shown. The stimuli either referred to real objects (i.e., *go* condition) or to meaningless ones (i.e., pseudo-words and scrambled images; *no-go* condition). In the *go* condition, participants responded significantly slower when nouns and pictures of graspable objects were presented. According to the authors, activating the motor system both when manually responding and when processing a graspable object comes with a cognitive cost, hence the slower response times. A similar effect was found in a previous study by Marino et al. (2014), who tested English native speakers, leading Buccino et al. (2017b) to conclude that motor response modulation was similar in L1 and in L2.

Dudschig et al. (2014) tested a similar effect in L1-German L2-English late bilinguals. In their adapted Stroop task, participants had to identify colors of the presented words using downward or upward motor responses. The presented words referred to

entities with a typical location (e.g., *star*, *root*) (Experiments 1, 2) or emotions (Experiment 3). The authors showed that responses were faster when words matched participants' motor responses (e.g., upward response with the word *star* or the word *happy*, experientially associated with "up") in both languages.

According to Dudschig et al. (2014), such facilitation could be due to (a) an automatic activation of L1 words and their experiential associations when processing L2 words or (b) a direct connection made during L2 learning to the sensorimotor experiences made during L1 learning. Even if the latter interpretation was favored due to the early onset of the embodiment effect, the former cannot be excluded, as the results by Vukovic and Williams (2014) suggest. In their study, 24 L1-Dutch L2-English bilinguals listened to English sentences implying physical distances (e.g., *On the plate in front of you, you can see a bone* vs. *On the plate at the far end of the table, you can see a bone*), with interlingual homophones (e.g., "bone," which in Dutch sounds like the word "boon" [beans]/bo:n/). After each sentence, a picture of the target object was presented to participants, in small or large dimensions. Large pictures were congruent to the sentences implying near distances and the small ones to those implying far distances. Participants were slower in judging if an object had been mentioned in the sentence previously heard if that object was a homophone in L1 with perceptual features congruent to the distance implied by the sentence. The authors argued that a perceptual simulation supports an early and parallel semantic processing in the two languages. Namely, bilinguals mentally simulate detailed perceptual features of L1 homophones while processing L2.

In their adapted Stroop task, Ahlberg et al. (2017) used the German spatial prepositions *auf* [on], *über* [above] and *unter* [under/below]. Participants, native or non-native German speakers – one non-native group with a similar use of spatial prepositions (i.e., English or Russian) and one non-native group with a dissimilar use of spatial prepositions (i.e., Turkish or Korean) – had to identify colors of the presented words using an upward or a downward hand movement. Results showed a different pattern of embodiment depending on L2-proficiency and on the corresponding use of the prepositions in the non-native groups' L1. However, all three groups (native, non-native similar and non-native dissimilar) were similarly affected by the Stroop task: responses were faster when the hand movement matched the spatial direction of the preposition. The authors concluded that processing a word in L2 does activate an experiential trace created in L1. This in turn corresponds to the first interpretation of Dudschig et al. (2014) and is in line with the results of Vukovic and Williams (2014), supporting the idea of a co-activation of L1 and L2. However, it should also be noted that a co-activation of L1 and L2 does not necessarily rule out the possibility of a direct, newly built connection between L2 words and the experiential representations.

Others have been less inclined to suggest that L2 was embodied, at least as strongly as L1. For example, Qian (2016) showed stronger embodiment effects in L1 than in L2. In her paper, she investigated the way the vertical spatial metaphor of the concept of "power" was processed in L1-Chinese L2-English speakers, half of them having high L2-English proficiency.

Participants had to judge if the nouns presented on the upper or lower part of the screen were related to “power” or not. Words associated with higher power were facilitated when presented in the upper part of the screen, whereas words associated with lower power were facilitated when presented in the lower part of the screen. This effect was, however, stronger in L1 than in L2, and was stronger in L2 for higher proficient L2 speakers. Note that some limitations of this study, both methodological and statistical (e.g., lack of detailed report) force us to consider its results with caution.

Still, a number of language studies, in which emotional valence of the stimuli was manipulated, have also observed differences in L1 and L2 affective processing, suggesting that the languages may be embodied to a different extent, especially in the case of late acquired L2 (Pavlenko, 2012). For example, Sheikh and Titone (2016), focusing on early stages of lexical processing, found L1-French L2-English speakers to be faster to process positive words than neutral words (first time reading passes), but not faster to read negative words than neutral ones. This was not the case in their previous work on L1 (Sheikh and Titone, 2013), suggesting, as raised by the authors, that negative words do not seem to be grounded in emotional experiences in L2. However, the concreteness advantage (sensorimotor grounding) in L1 was present for low frequent neutral words but not for emotional words (Sheikh and Titone, 2013), while in L2 it was present for both neutral and negative high frequent words (Sheikh and Titone, 2016). Moreover, results showed that L2 proficiency positively correlated with the concreteness advantage.

In sum, behavioral studies revealed that L2 is very likely embodied. Firm conclusions regarding the degree to which L2 is embodied remains to be clarified, as some studies report differences in L1 vs. L2 embodiment (Qian, 2016; Sheikh and Titone, 2016; Ahlberg et al., 2017) whilst others did not find such differences (Dudschig et al., 2014, Experiment 1), or did not perform direct statistical comparisons between languages (Bergen et al., 2010; Vukovic and Williams, 2014; Buccino et al., 2017b). In **Tables 1, 2**, we summarize the studies that have investigated these issues.

(Neuro-)Physiological Studies

To our knowledge, De Grauwe et al. (2014) were the first to conduct an fMRI study to investigate embodiment in L2. In a lexical decision task, highly proficient L1-German L2-Dutch and Dutch native speakers were presented with motor and non-motor cognate or non-cognate⁴ verbs in Dutch. Results showed a significantly stronger activation in motor and somatosensory areas for motor verbs, regardless of the cognate status of the verbs. This was the case for both language groups. De Grauwe and colleagues consequently suggested L2 representations to be rich enough to activate similar motor-related areas as L1. Note that as all participants were late highly proficient bilinguals, the impact of proficiency and AoA on the embodiment effect cannot be established beyond conjecture (De Grauwe et al., 2014).

⁴Cognates are words that share orthographic and/or phonologic features between languages (e.g., *nemen* in Dutch with *nehmen* in German [to take]).

In a similar vein, Xue et al. (2015) presented L1-Chinese L2-English participants with high (e.g., *crumb*) and low (e.g., *lace*) body-object interaction (BOI) English words. These words were imbedded in high (e.g., *you brush the small sticky crumb*) and low (e.g., *you wear a string of cotton lace*) sensorimotor contexts. Highly proficient L2-English participants judged sentence acceptability while ERPs time-locked to the onset of the high vs. low BOI words in rich and poor context were recorded. The results showed a marginal sensorimotor context effect reflected in ERP differences in both the P2 and N400 components. The authors suggested that this effect was related to differential activation of sensorimotor areas, based on observed differences in electrodes over the sensorimotor cortex.

Other studies including neurophysiological measures have also supported the notion that bilinguals' L2 is less embodied than L1. Vukovic and Shtyrov (2014), for example, examined mu-rhythm event-related desynchronization as an index of motor cortex activity in response to L1 and L2 abstract and action prime-probe verb pairs. Highly proficient L1-German L2-English speakers performed a passive reading task while an electroencephalogram was recorded. Analysis of motor-related EEG oscillations revealed that cortical motor activation was present in both L1 and L2 around 150 ms post-stimulus. Yet, L1 probe verbs elicited stronger sensorimotor brain activation than L2 probes. Foroni (2015) measured the strength of zygomatic muscle activation when participants read relevant (i.e., to the zygomatic muscle) affirmative and negative short sentences (e.g., *I am... or I am not... smiling*) and irrelevant ones (e.g., *I am... or I am not frowning*). Having *negative* sentences provided the authors with an alleged muscle relaxation condition, offering a way to further evaluate inhibition processes. Interestingly, the results showed stronger activation of the zygomatic muscle when participants read affirmative sentences, mimicking the results found in L1 (Foroni and Semin, 2013). Yet, the magnitude of the somatic activation was smaller in L2 than L1. Moreover, differently from L1 (Foroni and Semin, 2013), there was no relaxation of the relevant muscles when participants read negative sentences in L2. Therefore, embodiment in L2 was only partial.

These results are corroborated by those of Baumeister et al. (2017) on emotion and memory. Grounded in the idea that emotional words are better remembered than neutral ones, they recorded electromyography and skin conductance of 26 late L1-Spanish L2-English bilinguals during a categorization task of emotional and neutral words in both L1 and L2. A day later, participants went through a memory recognition task. Although their results were not decisive (i.e., marginally significant), there were some trends indicating that (a) there was a reduced, delayed and short-lived motor resonance in response to emotional words in L2, and that (b) a strong motor resonance would lead to better memorizing of emotional words.

Some studies on bilingualism and emotions (e.g., Harris et al., 2003; Harris, 2004; Caldwell-Harris, 2015; Hsu et al., 2015) have also suggested that L2 emotional words evoke less autonomic physiological response than L1 words, leading some authors to describe L2 as “disembodied” (for a review see Pavlenko, 2012, 2017). However, as Sheikh and Titone (2013) have pointed out, there might be a difference between *emotionally grounded* and

sensorimotor grounded concepts, difference which goes beyond the scope of this paper.

In sum, (neuro-)physiological data globally confirm findings from behavioral ones on L2 embodiment, independent of the techniques used. Some issues still remain unanswered though, especially those pertaining to the degree by which L2 is embodied and to the roles of AoA, proficiency and immersion (see **Tables 1, 2** for a summary of these studies).

Critical Synthesis

The role of the sensorimotor system in L2 language processing has not received much attention, yet we have tried to gather and collate the few studies specifically focused on this issue. Crucially, all these studies show an embodiment effect during the lexico-semantic processing of L2 (see **Table 2**), independently of the techniques used (behavioral or neurophysiological) or of the specific aim of the study in question.

Interestingly, eight out of the twelve studies reported in this review *statistically* compared the degree of L2 vs. L1 embodiment (see **Tables 1, 2** for a summary), and only two of them concluded a similar embodiment for both languages (De Grauwe et al., 2014; Dudschig et al., 2014). However, in the latter two studies, the extent of true similarity would need further investigation. For example, Dudschig et al. (2014) reported a slightly stronger significance of embodiment effect in L1 vs. L2, without delving into it in the discussion, and De Grauwe et al. (2014) found different patterns in sensorimotor activation between L1 and L2, which they explained in terms of methodological parameters. All the other studies discussed in this review report that L2 is differently embodied than L1, usually expressed as a *lower degree* (Vukovic and Shtyrov, 2014; Foroni, 2015; Qian, 2016; Baumeister et al., 2017) of embodiment in L2 or as a *different pattern* (Sheikh and Titone, 2016; Ahlberg et al., 2017) of embodiment. Such a difference may be explained by different factors discussed hereafter.

Several studies suggest an influence of participants' L2 proficiency on the degree of L2 embodiment. In terms of the RHM model (Kroll and Stewart, 1994), and as suggested by others (e.g., Qian, 2016), this could be explained by an asymmetry in the strength of the connections between words and their representations in the two languages, mainly characterized by stronger links, and hence faster access to meaning, in L1. In contrast, access to L2 representations would require mediation via L1, especially in case of low L2 proficiency. This entails a later sensorimotor involvement when L2 proficiency is low compared to when it is high, or compared to L1. Such differences in the degree of L2 embodiment would also be in line with the BIA+ model (Dijkstra and van Heuven, 2002) assuming later semantic access when L2 proficiency is low. However, none of the studies presented can actually reach a definite conclusion as to the role of proficiency, and this for three main reasons. First, L2 proficiency was not always thoroughly assessed, if assessed at all. To provide us with relevant insight into the issues discussed so far, we believe that L2 proficiency should always be assessed, whether it be on objective measures such as receptive (e.g., DIALANG, Zhang and Thompson, 2004), and productive vocabulary (e.g., Productive Vocabulary Levels Test,

Laufer and Nation, 1999), and/or subjective ratings from questionnaires including self-evaluation and language background (e.g., LEAPQ, Marian et al., 2007). Second, L2 proficiency was never actually specifically manipulated (except in Qian, 2016, without thorough proficiency assessment). Third and finally, participants' L2 general proficiency could not always be reflected in the actual lexico-semantic knowledge of the stimuli in the experiment, therefore raising the need to add task-specific measurements of proficiency, as was done by Bergen et al. (2010), who administered a passive lexical knowledge test.

One could further argue that even if proficiency was to be carefully assessed, any embodiment effect could also be accounted for by factors such as exposure to L2 and/or AoA. If the degree of embodiment of L2 depends on the degree to which L1 and L2 share their semantic representations, some models (e.g., Silverberg and Samuel, 2004) would actually assume a common semantic system between languages only in the case of early AoA. Therefore, L2 lexico-semantic processing would involve sensorimotor areas to the same degree as L1 lexico-semantic processing only in case of an early acquired L2. Exposure and AoA have never been manipulated in bilingual studies on embodiment, allegedly the former because it may be highly interrelated to proficiency and the latter because it is usually considered to be less associated with semantic processing. This is rather unfortunate, as representations have been shown to be modulated by exposure when proficiency was kept constant (e.g., Perani et al., 2003), even after a short period (e.g., Dahl and Vulchanova, 2014). Not considering AoA may also be problematic, as AoA could show different effects depending on the nature of L2 learning. Namely, early L2 AoA has been associated with implicit L2 learning, which takes place in a naturalistic setting via sensorimotor experiences, while late L2 AoA has been associated with explicit learning, taking place in the setting of a traditional classroom via amodal instructions. Some studies contrasting different types of L2 learning have been mainly interested in learning and memory performances (e.g., Zimmer, 2001; Repetto et al., 2017; García-Gómez and Macizo, 2018). Other studies have tried to untie the type of learning from AoA. For example, independent of the learning setting, structural changes have been observed in the left inferior parietal cortex, and differences in these changes have been attributed to AoA (Stein et al., 2014).

In fact, the importance of the type of learning for L2 embodiment may be illustrated by studies which show a rapid association between motor areas' activation, or excitability, and novel labels attributed to actions or tools (e.g., Liuzzi et al., 2010; Fargier et al., 2012; Branscheidt et al., 2017a; Bechtold et al., 2018 in elderly). These studies showed embodiment effects in newly formed L2-like representations, also when experiential traces were not transferred from L1 to L2 (Fargier et al., 2012; Öttl et al., 2017; Bechtold et al., 2018). As such, these studies document the influence of exposure, AoA, and type of learning on grounding language in bodily experiences. Interestingly, and future research on these effects taking a lifespan perspective should consider this, language-induced motor activity in the brain has been shown to change with training (Fargier et al., 2012), and seems to be different between children and adult (Dekker et al., 2014),

TABLE 1 | Methodological description of the reviewed studies.

Authors	Type of study	L1	L2	N for analyzed data	Proficiency	Immersion	AOA	Task	Stimuli
Bergen et al., 2010	Behavioral	Experiment 1: L1 English; Experiment 4: L1 various, L2 English	English	39 (experiment 1); 35 (experiment 4)	General: High (students with a min TOEFL score). Task-specific: passive lexical knowledge test	Not reported	Not reported	Forced choice matching task	Verbs (written and imaged)
Buccino et al., 2017b	Behavioral	Italian	English	23	High, C1 (CEFR)	Not reported	Not reported	Go-no go paradigm	Photos and nouns of graspable and non-graspable objects; pseudowords and scrambled images
Dudschig et al., 2014	Behavioral (Experiment 1)	German	English	20	Not reported	Not immersed. Participants never lived in English speaking country	Late bilinguals: 11–13 y.o.	Modified Stroop (response to the words' ink color with an upward or downward arm movement)	Written L1 and L2 words referring to entities with a typical location (e.g., star, mole)
Vukovic and Williams, 2014	Behavioral (Experiment 2)	German	English	20	Not reported	Not immersed. Participants never lived in English speaking country	Late bilinguals: 11–13 y.o.	Written L2 words referring to entities with a typical location (e.g., star, mole)	Written L2 words referring to an emotion (e.g., happy, sad)
	Behavioral (Experiment 3)	German	English	20	Not reported	Not immersed. Participants never lived in English speaking country	Late bilinguals: 11–13 y.o.	Judge whether presented pictures depicted something mentioned in a previously heard sentence	English sentence-picture pairs using interlingual English–Dutch homophones (e.g., “bone,” which in Dutch sounds like the word “boon” [beans]/bo:n/) occurring in the final position.
Ahlberg et al., 2017	Behavioral	Group 1: German; group 2: Russian and English; group 3: Turkish and Korean	German	Group 1: 47; group 2: 45; group 3: 42	From A2 to C2, self-reported CEFR levels	Immersed. All participants were students or employees in Germany	Group 2: 8–28 y.o.; group 3: 3–26 y.o.	Modified Stroop (response to the words' ink color with an upward or downward arm movement)	Spatial prepositions: “auf”, “unter” and “über” and the filler “ab”
Qian, 2016	Behavioral	Chinese	English	30 (15 per group)	High (group A) and low (group B); Group A has passed the test for English Majors Band 4 (TEM-4), while group B has no certificate of College English Test Band-4 (CET-4)	All participants lived in China. No immersion reported. Group A students are majoring in English	Not reported	Semantic judgment	Written high vs. low power nouns in L1 and L2

(Continued)

TABLE 1 | Continued

Authors	Type of study	L1	L2	N for analyzed data	Proficiency	Immersion	AOA	Task	Stimuli
Sheikh and Titone, 2016	Behavioral (eye-movement measures)	French	English	34	Self-reported measures: modified LEAP-Q; Marian et al., 2007, with a 7-point scale assessing the amount of the L2 use in different contexts, and the percentage of time exposed to L2	Self-reported measures: modified LEAP-Q; Marian et al., 2007, with a 7-point scale assessing the amount of the L2 use in different contexts, and the percentage of time exposed to L2	Not reported	Natural sentence reading	Semantically neutral sentences in which embedded L2 target words varied on emotional valence (negative, neutral and positive), frequency and concreteness
De Grauwe et al., 2014	Neurophysiological (fMRI)	Group 1: Dutch group 2: German	Dutch	Group 1: 20 group 2: 18	High (min. 67.5% accuracy in an online lexical decision task, LexTALE)	At least 1.5 years living in the Netherlands plus regular usage of L2	Late: $M = 19.94$, $SD = 2.88$	Lexical decision task	Written motor- and non-motor cognate and non-cognate verbs/pseudowords
Xue et al., 2015	Neurophysiological (EEG)	Chinese	English	17	High (near-balanced), assessed with a self-rating scale, and all passed a national test for English majors, similar to TOEFL	Students majoring in English, but living in China	Not reported	Sentence acceptability judgment	Written high and low BOI words embedded in segmented sentences characterized by rich and poor sensorimotor context
Vukovic and Shtyrov, 2014	Neurophysiological (EEG)	German	English	18	High, measured by self-rated proficiency scores (7-point Likert scale, $M = 6.10$, $SD = 0.70$) and scores on the LexTALE (Lernhöfer and Broersma, 2012) English vocabulary test ($M = 81.82$, $SD = 22.34$)	Not reported. Participants started learning English as part of formal education in Germany	Late (mean 10:19 y.o., $SD = 2.19$)	Passive reading	Written L1 and L2 action and abstract prime-probe verb pairs
Foroni, 2015	Neurophysiological (EMG)	Dutch	English	26	Good fluency self-reported	Not reported	Late (after 12 y.o.)	Passive exposure to affirmative and negative sentences followed by a simple classification of arrow directions (left or right)	Written L2 affirmative and negative sentences either relevant (e.g., <i>I am smiling</i>) or irrelevant (e.g., <i>I am frowning</i>) to the target muscle under examination (zygomatic major)
Baumeister et al., 2017	Neurophysiological (EMG)	English or Spanish	Spanish or English	26	Advance level self assessed from 1 to 10 (LEAP-Q; Marian et al., 2007) and letter fluency task	At least 12 months in their L2 speaking country	Not clearly reported. Most of them have a late AoA (started to learn L2 in a classroom setting)	Facial muscle EMG activity and SC responses were obtained during the encoding phase of a classical memory task, in which participants performed a categorization task, which required them to categorize words into "associated to emotion" or "not associated to emotion."	Written L1 and L2 emotion-laden words

TABLE 2 | Theoretical description of the reviewed studies.

	Authors	Aim of the study	Comparison	Embodiment effect	Conclusions on L2 embodiment	L2 vs. L1 embodiment
Behavioral	Bergen et al., 2010	To investigate if (1) action and language understanding use the same motor circuitry and if (2) this motor activation plays a functional role in language understanding	Descriptive comparison between L1 and L2 and between subjects	Slower RTs when the image and verb shared an effector, than when they did not share an effector	<ul style="list-style-type: none"> • "Like native speakers, these nonnative speakers relied on motor structure activation to understand words" • "Language proficiency (as measured by overall accuracy in the main experiment) correlated positively with effect size" 	L2 is embodied/ No statistical comparison with L1
	Buccino et al., 2017b	To test if L2 speakers showed the same kind of modulation of motor responses as participants in a previous experiment (Marino et al., 2014)	Descriptive comparison between L1 and L2 and between subjects (previous study Marino et al., 2014)	Slower RTs during the processing of graspable items as compared to non-graspable ones	<ul style="list-style-type: none"> • "Fluent speakers of English as L2, showed the same kind of modulation of motor responses as participants in a previous experiment (Marino et al., 2014), where the same kind of stimuli were presented in their L1" 	L2 is embodied/ No statistical comparison with L1
	Dudschig et al., 2014	To investigate basic associations between L2 and the sensorimotor system	Statistical comparison between L1 and L2 within subjects	Compatibility effect, reflected in an interaction between the implicit location of words and the direction of the response movement	<ul style="list-style-type: none"> • "L2 automatically activated motor responses similar to L1 even when L2 was acquired rather late in life (age > 11)" • "Reactivation of experiential traces is not limited to L1 processing" 	No difference between L2 and L1 embodiment
			None	Compatibility effect, reflected in an interaction between the implicit location of words and the direction of the response movement		L2 is embodied/ No statistical comparison with L1
			None	Compatibility effect, reflected in an interaction between emotion words (e.g., happy, sad) and the direction of the response movement		L2 is embodied/ No statistical comparison with L1
	Vukovic and Williams, 2014	To investigate the automaticity of embodied simulatory processes in language comprehension, exploiting the fact that both languages are activated during comprehension in bilinguals	None	Slower RTs to target stimuli in the homophone condition, suggesting activation of both languages which, in turn, led to greater competition between alternative meanings	<ul style="list-style-type: none"> • "Perceptual simulation of L1 meaning occurs during L2 sentence processing" 	L2 is embodied/ No statistical comparison with L1
	Ahlberg et al., 2017	To compare embodiment effects related to the processing of spatial prepositions in German native speakers with the embodiment effects potentially observed in different groups of L2 learners	Statistical comparison within language (German as L1 and as L2) and between subjects	Compatibility effect, reflected in an interaction between the meaning of the spatial prepositions and the direction of the response movement	<ul style="list-style-type: none"> • "Reactivation of experiential traces plays an important role not only in L1 processing but also in the processing of L2, even if this [L2] is quite dissimilar to the participants' native language" . . . "We found clear processing differences between this group [dissimilar-L1 group] and the other two groups" • "We found [...] also clear differences between speakers of different L1s" + "L2 language proficiency modulate the embodiment effect" 	L2 is embodied, but differently from L1

(Continued)

TABLE 2 | Continued

Authors	Aim of the study	Comparison	Embodiment effect	Conclusions on L2 embodiment	L2 vs. L1 embodiment
Qian, 2016	Investigate if (1) Chinese English learners with different proficiency levels activate vertical spatial metaphors in processing power words, and if (2) L2 proficiency has an impact on the mental representations in processing power words	Statistical comparison between L1 and L2 within participants, and statistical comparison within L2 between participants (high vs. low proficient)	Faster RTs for power words presented in the upper vs. lower part of the screen	<ul style="list-style-type: none"> • “Chinese English learners tend to show stronger mental representation abilities to their first language than their second language” • “High-level second language learners have stronger representation ability than low-level learners in understanding English words of power” • “For RT, there’s an interaction between power words and L2 proficiency. [...] Learners’ second language proficiency had an impact on their mental representation ability.” 	L2 is embodied, but differently from L1
Sheikh and Titone, 2016	To investigate if bilingual readers (1) exhibit L2 emotional word processing effects, (2) show facilitation of negative words by concreteness, like neutral words, and if (3) differences in L2 proficiency predict facilitation by concreteness and frequency, but not emotionality	Statistical comparison within L2 across participants with varying L2 proficiency and statistical comparison within language (English as L1 and as L2) between participants (previous study Sheikh and Titone, 2013)	Shorter gaze duration (concreteness facilitation) for negative and neutral low-frequency words, but only at high levels of proficiency	<ul style="list-style-type: none"> • “Previous work on L1 embodiment indicates that a concreteness advantage, where observed, is diagnostic of emotional neutrality because it does not occur for emotionally charged words (Sheikh and Titone, 2013).” • “Bilinguals have emotionally disembodied negative words during L2 reading, and that these words are instead grounded in sensorimotor experiences like neutral words.” • “L2 proficiency predicts concreteness advantages but not emotional advantages during natural reading.” 	L2 is embodied, but differently from L1
Neurophysiological De Grauwe et al., 2014	(Among others) to investigate if L2 speakers show embodiment effects like L1 speakers	Statistical comparison between Dutch as L1 and Dutch as L2 between participants	Motor verbs yielded significantly higher activation in sensorimotor ROIs than non-motor verbs	<ul style="list-style-type: none"> • “Both groups [L1 and L2 speakers] displayed higher activation for motor than for non-motor verbs in motor and somatosensory brain areas. These activations were evident with both cognate and non-cognate verbs, indicating that they were not due to transfer from the native language of the L2 speakers.” 	No difference between L2 and L1 embodiment
Xue et al., 2015	To investigate if (1) richer sensorimotor context lead to increased sensitivity to the anticipated sensorimotor consequences of BOI effect, and if (2) the context and use of BOI word processing activate sensory- and motor-related brain areas	None	Sensorimotor context had an impact on the processing of BOI words that this modulation did activate related sensorimotor areas	<ul style="list-style-type: none"> • “Action- and perception-related brain areas for L2 words are activated, indicating that the semantic representations for L2 learners are rich enough for the sensorimotor-related activation.” 	L2 is embodied/No comparison with L1

(Continued)

TABLE 2 | Continued

Authors	Aim of the study	Comparison	Embodiment effect	Conclusions on L2 embodiment	L2 vs. L1 embodiment
Vukovic and Shtyrov, 2014	To investigate if the motor cortex of bilingual subjects shows differential involvement in processing action semantics of native and non-native words	Statistical comparison between L1 and L2 within participants	Action related words elicited increased motor brain activity reflected in stronger ERD	<ul style="list-style-type: none"> “Processing of action verbs was accompanied by early motor activation for probe stimuli in both languages of bilingual subjects, as reflected in desynchronization of the EEG mu rhythm. Furthermore, at the level of sensor data and source activation clusters, we observed that this motor activation is stronger in the L1, likely due to highly integrated action-perception circuits formed as a result of rich linguistic experience.” At the sensor level, L1 probes elicited a larger desynchronization in the right-hemisphere cluster than L2 probes did; at the source level, L1 action primes, unlike L2 ones, produce more ERD. 	L2 is embodied, but differently from L1
Foroni, 2015	To examine if the processing of affirmative and negative sentences in L2 relies on the same somatic bases as that of L1 (Foroni and Semin, 2009, 2013)	Statistical comparison between L1 and L2 between participants (across studies, L1 data from Foroni and Semin, 2013)	When participants read affirmative sentences (e.g., ‘I am smiling’) the relevant muscle (i.e., zygomatic major) activates; however, when they read negative sentences (e.g., ‘I am not smiling’), the relevant muscles inhibited only in L1	<ul style="list-style-type: none"> “The magnitude of the somatic reaction in L2 is smaller than the one reported for L1” “The weaker magnitude of the somatic simulation for L2 compared to the one reported for L1 is generally in line with the argument that the different socialization histories of L1 and L2 are reflected in different degree of embodiment” For affirmative relevant sentences: L1 shows larger activation compared to L2 at 600 and 1000 ms period with the 800ms period showing a marginal significant difference. For negative relevant sentences, significant relaxation/inhibition in L1 and no effect in L2. 	L2 is embodied, but differently from L1
Baumeister et al., 2017	To investigate the link between embodied processes and memory for emotional content within the frame of L1 and L2 processing	Statistical comparison between L1 and L2 within participants	Differential activation in the zygomaticus and corrugator muscle in response to happy vs. angry words	<ul style="list-style-type: none"> “The overall results of the EMG and SC recordings suggested some reductions and differences in embodied simulations of emotional L2 words in comparison with emotional L1 words” “The difference between L1 and L2 processing became particularly clear in the corrugator muscle, which showed typical response patterns to emotional L1 words but no detectable responses to emotional L2 words.” “The present results complement and extend Foroni’s results showing a significant difference for the corrugator muscle and, thus, together they support an embodiment account of emotion processing and a reduced embodiment in L2.” Result indicates a later onset and shorter duration of specifiable zygomaticus activity in response to happy vs. angry words in L2 compared with L1 	L2 is embodied, but differently from L2

yet already present in young children (e.g., James and Swain, 2011; see also Inkster et al., 2016). These issues have been well documented (e.g., Macedonia, 2014; Wellsby and Pexman, 2014; Macedonia and Mueller, 2016).

Another factor that could account for differences between L1 vs. L2 embodiment is the linguistic distance between languages, which refers to the extent of similarity between the languages and which has previously been shown to play a role in bilingual language processing (e.g., Lindgren and Muñoz, 2013; Abutalebi et al., 2015; Ghazi-Saidi and Ansaldo, 2017). This factor is usually studied in relation to the ease of learning a second language (e.g., Butler, 2012), or in relation to the phonology and morpho-syntax of languages (e.g., Llama et al., 2010; Zawiszewski et al., 2011). Studies on the influence of linguistic distance on embodiment remain scarce and languages have not been always chosen in a systematic way. For example, some have compared embodiment in languages that are both Germanic (Vukovic and Shtyrov, 2014; Foroni, 2015), others compared a Germanic language to an Italic one (Sheikh and Titone, 2016; Baumeister et al., 2017), and Qian (2016) compared two different language families (i.e., Sino-Tibetan and Indo-European). Essentially, linguistic distance could act as a catalyst for embodiment similarity between L1 and L2. To the best of our knowledge, only one study addressed this issue (i.e., Ahlberg et al., 2017), and found little effect of linguistic distance. In a nutshell, Ahlberg et al. (2017) found similar embodiment effects in L2 (German) for two non-native groups, irrelevant of the linguistic distance between L1 and L2 (i.e., whether or not L1 linguistic properties could easily match to L2). Clearly, more research needs to be carried out to reach definite conclusions.

This issue is nonetheless relevant, especially in studies that (a) involve words with a special status (e.g., cognates, as in De Grauwe et al., 2014; or false friends, as in Degani et al., 2018, and Persici et al., 2019), (b) involve manipulating linguistic properties that differ across languages (e.g., the meaning of spatial prepositions, as in Ahlberg et al., 2017; or the perspective implied by the use of personal pronouns, as in Papeo et al., 2011) or (c) involve an experimental design in which the two languages are intermixed in the same block event (e.g., semantic priming driven by phonological properties, as in Vukovic and Williams, 2014; Degani et al., 2018).

Others have stressed the timing of the motor system involvement as an explanatory factor for the difference between L1 and L2 embodiment. Differences both in the onset of the motor resonance and its duration have been reported by Foroni (2015) and Baumeister et al. (2017). Specifically, their experiments showed that L2 motor resonance had a later onset and shorter duration compared to L1. Latency shifts have previously been associated with delayed lexico-semantic processing for L2 compared to L1 in several neurophysiological studies (e.g., Moreno and Kutas, 2005; Leonard et al., 2010; Newman et al., 2011), in line with the bilingual language models suggesting faster access to meaning in L1, as discussed earlier.

Arguably, these potential explanatory factors – all legitimate – raise an important issue, as to the stages of cognitive processing under investigation. Accordingly, any endeavor to investigate embodiment in L2 should always be very explicit as to which

stage of processing is under investigation. This is crucial, as the majority of the studies on this topic used tasks which allegedly access early stages of lexical processing (e.g., a Stroop task or a lexical decision task, where the access to meaning is not necessary; Dudschig et al., 2014; Ahlberg et al., 2017), while others used tasks which require deep semantic processing (e.g., a semantic judgment or a picture-word matching task; e.g., Vukovic and Williams, 2014; Xue et al., 2015; Qian, 2016). As differences in embodiment related to the depth of semantic processing have been shown in L1 (e.g., Willems et al., 2009; Vukovic et al., 2017), we would further argue the motor circuit recruitment to be different between L1 and L2 depending on the task used – consequently the stage of processing accessed – in the experiment.

Importantly, all explanatory factors – to differences between L1 and L2 embodiment – presented so far have been based on studies on *language-to-motor* effects. A more complete (or even different) picture of the interaction between the sensorimotor system and lexico-semantic processing may stem from also examining *motor-to-language* effects. This may be crucial, as we do know, from studies on monolinguals, that experimental manipulations of the sensorimotor system can affect lexico-semantic processing. Sensorimotor system manipulations have been as diverse as motor training (e.g., in healthy Glenberg et al., 2008; Locatelli et al., 2012; in experts Beilock et al., 2008; or with dyslexic children Trevisan et al., 2017), motor limitation (e.g., Bidet-Ildei et al., 2017), or motor brain area stimulation (Willems et al., 2011; Tremblay et al., 2012; Vukovic et al., 2017; Gijssels et al., 2018). To the best of our knowledge, no study has directly assessed *motor-to-language* effects in healthy bilinguals, linking the sensorimotor system and lexico-semantic processing. Interventions on the motor system may help language processing, as much as language-based interventions may contribute to motor improvements, both in L2 and L1. More generally, and this is the focus of the next section, we believe that studies on L2 embodiment may serve also clinical purposes, although this has been only rarely recognized.

STUDIES ON L2 EMBODIMENT SERVING CLINICAL PURPOSES

No clinical study has apparently explicitly linked the sensorimotor system to L2 lexico-semantic processing. Nonetheless, some studies on bilingual patients with motor impairment did explore motor-language interactions, yet with somehow different purposes (e.g., syntactic impairment). In the next section, we discuss these studies and corollary hypotheses related to lexico-semantic processing. In the following section, we present some clinical rehabilitation studies – in L1 – that could be interpreted in terms of embodiment (e.g., language-action therapies in aphasic patients) and then extend the discussion to L2, and bilingual rehabilitation outcomes.

Motor-Language Interactions

Clinical studies on the interaction between motor and L2 language systems have been scarce, yet could document the

modulation of motor impairment on L2 processing as well as the impact of L2 impairment on sensorimotor systems.

In Section “Embodiment Predictions for L2 and Their Impact on Language Models” we discussed the idea that L2 lexico-semantic representations should be less grounded in the sensorimotor system – the motor cortex – if L2 is acquired through late explicit learning. This is reminiscent of the Procedural/Declarative model of language acquisition (Ullman, 2001), which distinguishes between procedural memory – which underlies implicit linguistic competences – and declarative memory – which underlies explicit linguistic competences –. The former is implemented in fronto-basal ganglia circuits, whilst the latter is implemented in bilateral medial and temporoparietal structures. In light of this model, Zanini et al. (2004, 2010) and Johari et al. (2013), for example, discussed how implicit grammatical language processing in L1 is more impaired than explicit grammatical language processing in a late L2 in Parkinson’s disease, as one would expect from a disease characterized by an impairment in the fronto-basal ganglia loops. In Johari et al. (2013), Parkinsonian patients did more error in L1 (implicit learning) than in L2 (explicit learning) in all the three administered syntactic tests from the Bilingual Aphasia Test, whilst this was the case only in one subtest for healthy controls. Importantly, these deficits were not correlated to other cognitive measures such as the Mini Mental State Examination, the Wisconsin Card Sorting Test and the Colored Raven Progressive Matrices, illustrating their specific linguistic focus. Similarly, Zanini et al.’s (2004, 2010) Parkinsonian patients showed deficits in syntactic processing and more phonological and morpho-syntactic errors in L1 than in L2, whilst healthy controls had fewer errors in L1 than in L2.

Whilst proficiency, exposure to L2 and AoA were not always carefully considered in studies on healthy participants, these factors were more thoroughly reported in Zanini et al. (2004, 2010) and Johari et al. (2013). In fact, in these studies, both healthy controls and patients (a) were proficient in L2 (based on the number of years and the context of usage), (b) were exposed to L2 on a daily basis, and (c) had acquired L2 late (at 6 years old at school). Participants in Johari et al. (2013) were highly proficient L2 speakers, and L2 was also their dominant language (used every day). Even if not specifically manipulated or formally assessed, Johari and colleagues argued that high L2 proficiency could explain worse performance in L2 in patients vs. controls, whilst the performance in L2 was not affected in lower proficient speakers in Zanini et al. (2004, 2010). The authors suggested that in case of higher proficiency, L2 is more likely to be processed partly implicitly, as L1, hence relying on procedural as well as declarative memory (Hamrick et al., 2018). Clinical studies specifically focusing on L2 lexico-semantic and sensorimotor systems (and their related brain areas) are needed to better understand procedural and declarative language influences on the motor network (and vice-versa). In fact, studies on monolingual patients showed that semantic deficits (declarative knowledge) affect more severely action-related than non-action-related stimuli in Parkinson’s disease (e.g., Cardona et al., 2013; Bocanegra et al., 2015; Gallese and Cuccio, 2018), which does not seem to be predicted by

the Procedural/Declarative model (see also Druks and Weekes, 2013). Note that Zanini et al. (2010) did suggest grammatical properties to be accessed during lexical retrieval, and therefore hinting at the idea that lexico-semantic knowledge may be connected to morpho-syntactic properties of language. As such, disentangling syntactic from lexico-semantic processes might not always be possible (e.g., Zwaan et al., 2010; Sell and Kaschak, 2011; Ahlberg et al., 2017).

Data on bilingual Parkinsonian patients also illustrate the Disrupted Motor Grounding Hypothesis (DMGH; Birba et al., 2017), based on neural reuse theories (neural exploitation hypothesis, Gallese and Lakoff, 2005; Gallese, 2008; shared circuit model, Hurley and Chater, 2005; Hurley, 2008; neuronal recycling hypothesis, Dehaene and Cohen, 2007; massive redeployment hypothesis, Anderson, 2007a,b, see Anderson, 2010 for a review). These suggest that low-level neural circuits can be exploited, recycled, and redeployed for other cognitive functions than their original ones. Based on this idea, the DMGH suggests that impairment in the network responsible for sequencing motor information can disrupt the functionally corresponding higher-level mechanism of sequencing words (i.e., syntactic processing).

Importantly, and central to the present paper, the DMGH also predicts lexico-semantic deficits in motor-related disorders. According to the DMGH, action-related meanings, in a somatotopic manner, are mapped onto motor circuits. Accordingly, semantically processing action words and sentences, as well as integrating verbal and motor information, should also be impaired in Parkinsonian patients, which seems to be the case (Boulenger et al., 2008; Cardona et al., 2013; Fernandino et al., 2013a; García and Ibáñez, 2014; Bocanegra et al., 2015; García et al., 2016; Buccino et al., 2017a; Gallese and Cuccio, 2018; see also Bak, 2013 for a review including other motor neuron diseases). For example, in Boulenger et al. (2008), masked priming effects for action words were almost absent in Parkinsonian patients deprived of dopaminergic treatment, whilst they were present – as healthy controls – when they were on Levodopa. The author concluded that their results constituted compelling evidence that lexico-semantic processing depended on the integrity of the motor system (brought by the medication for Parkinsonian patients). Noteworthy, all patients in the studies of Zanini et al. (2004, 2010) and Johari et al. (2013) were on Levodopa or other dopaminergic drugs, but this condition was not enough to restore the intrinsic impairment in syntactic processing. As pointed out by Boulenger et al. (2008) reaction times or error rates for action verbs in their study were not differently affected by the motor impairment or by the dopaminergic treatment. Whether lexico-semantic impairment of action-related meanings and of other verbal and motor information integration is expected in L2 is yet to be examined. At least in L1 patients with basal ganglia impairment, who typically show frontostriatal atrophy, difficulties in motion-related verbal expressions seem to be detectable before the appearance of clinical symptoms (Birba et al., 2017). As such, linguistic diagnostic tasks may help identify Parkinson patients well before the clinical manifestation of the disease (Cardona et al., 2013; García and Ibáñez, 2014; García et al., 2017, 2018).

These tasks may also help to identify and stage pre-symptomatic Huntington disease patients (Kargieman et al., 2014).

Questions remain as to the use of linguistic diagnostic tasks in L2. At this point, there is no data to evaluate patients' sensitivity to L2 tasks that evaluate the processing ease of motion-related verbal expressions. Depending on the grounding of L2, a simple use of a diagnostic L1 task (yet to be generated) may not be adequate. Factors such as AoA and language competence may be critical, together with the presence of emotionally charged content, which might be perceived very differently depending on the language in use (i.e., L1 or L2, see Sheikh and Titone, 2016). Still, the few studies with bilingual Parkinsonian patients suggest that L2 linguistic diagnostic tasks could mimic L1 tasks, even for distant languages. Similar patterns of impairment in each language have been found in speakers of distant languages (e.g., two Indo-European languages in Zanini et al., 2010, and one Indo-European L1 and the other Altaic-Turkic L2 language in Johari et al., 2013). As previously suggested, the extent of language distance and its impact on these issues are yet to be thoroughly examined.

In sum, actual evidence on *motor-to-language* oriented clinical studies show four important findings. First, motor impairments impact lexico-semantic processing of motor related stimuli in L1 (e.g., Bak, 2013; Cardona et al., 2013; Fernandino et al., 2013a; Bocanegra et al., 2015). Second, motor impairments may impact morpho-syntactic processing in L2 (Zanini et al., 2004, 2010; Johari et al., 2013). Third, motor-related interventions could modulate language performances (Boulenger et al., 2008). Fourth and finally, all the factors discussed in the previous sections of this paper (i.e., proficiency, AoA, exposure, distance between languages, type of exposure) may influence the degree of language impairment due to motor-related diseases (Johari et al., 2013).

Although *motor-to-language* clinical studies in L2 may be scarce, there seems to be none on *language-to-motor* effects in L2. In other words, the impact of L2 lexico-semantic processing on motor system has yet to be examined in brain-damaged populations. In monolinguals, some studies did look at the co-occurrence of language and motor impairment in developmental disorders (e.g., Hill, 2001; Sanjeevan et al., 2015) or brain-damaged patients (Desai et al., 2015; for a review see Anderlini et al., 2019).

We believe that, however, weak the *language-to-motor* effects might be in L2 and unhealthy populations, they deserve some empirical attention, especially as they might give rise to linguistic markers of motor impairment.

Language-Motor Rehabilitation

As mentioned earlier, experimental manipulations of language in healthy monolinguals (e.g., Aziz-Zadeh et al., 2006; Boulenger et al., 2009; Alemanno et al., 2012; Ghio et al., 2018) and bilinguals (see section Behavioral Studies and (Neuro-)physiological Studies) have been shown to impact the motor system. Conversely, experimental manipulations of the motor system in healthy monolinguals have been shown to impact lexico-semantic processing (e.g., Beilock et al., 2008; Glenberg et al., 2008; Willems et al., 2009; Locatelli et al., 2012; Tremblay et al., 2012; Bidet-Ildei et al., 2017; Vukovic et al., 2017;

Gijssels et al., 2018). Moreover, experimental manipulations of the motor system in healthy bilinguals has been shown to impact visual perception of motor speech movements (e.g., Swaminathan et al., 2013). Importantly, no study has investigated the impact of experimental manipulations of the motor system on lexico-semantic processing in L2. Moving toward clinical studies, others examined the impact of experimental manipulations of the motor system in monolingual patients on lexico-semantic processing (e.g., dopaminergic treatment in Boulenger et al., 2008; motor training with dyslexic children in Trevisan et al., 2017).

With respect to neuromodulation interventions, transcranial direct current stimulation (tDCS) and TMS of the motor cortex of aphasic patients is of particular interest. While brain stimulation is increasingly being tested as promising auxiliary therapeutic tools in patients with aphasia, results have so far been inconsistent, the activation of different brain regions showing very different efficacy (Arévalo et al., 2007; Marangolo et al., 2016; see also Elsner et al., 2013; Lefaucheur et al., 2017 for reviews). The stimulation of the motor cortex is especially interesting considering that this region is easily located and it is often spared in aphasic patients (Branscheidt et al., 2017b; Dreyer and Pulvermüller, 2018). Recently, Branscheidt et al. (2017b) showed a specific role of the motor-cortex in accessing lexical-semantic content. Similarly, Meinzer et al. (2016) showed improved naming abilities after 2 weeks of concurrent speech and language therapy and left motor cortex stimulation. However, while these studies investigated effects of neuromodulation techniques on L1 processing, this question has not yet been addressed with bilingual patients. To the best of our knowledge, no clinical study has directly investigated the interaction between sensorimotor areas and L2.

With respect to behavioral interventions, we believe several methods to be relevant. Therapists can choose, for example, to reinforce the damaged language-specific neural network by training the specific language impairment or to work on a more general cognitive-control network reinforcing executive functions, or, in light of the studies on embodiment mentioned so far, strengthen the sensorimotor circuit. Several speech and language therapeutic approaches that are based on the interaction between the motor and the language systems, as in embodiment theories, have in fact shown promising results (e.g., Semantic Feature Analysis therapy, Boyle and Coelho, 1995; gestures production therapies, Krauss, 1998; Goldin-Meadow et al., 2001; Rose, 2006; Rose et al., 2013; Action Observation Therapy, Marangolo et al., 2010; language-action therapies, Difrancesco et al., 2012; Stahl et al., 2016). As an example, a motor recovery therapy based on the mirror neuron system, commonly called the Action Observation Therapy, has already been extended to the domain of aphasia. Marangolo et al. (2010) showed that after therapy, four non-fluent chronic lexico-phonological impaired aphasic patients improved in lexical retrieval as a result of both "action observation" therapy and "action observation and execution" therapy. Importantly, their improvement was still evident 2 months after the treatment. The authors suggested that the sensory-motor representations, activated by observing a performed action, served as input at the lexical level and facilitated word retrieval (Marangolo et al., 2010;

Bonifazi et al., 2013). However, one other study showed no improvement in two aphasic patients with the same type of therapy, which was attributed to differences in the cognitive and linguistic profiles of the patients (Routhier et al., 2015). Nonetheless, Gili et al. (2017) – using fMRI – recently confirmed Marangolo et al.'s (2010) hypothesis by showing a sensorimotor recruitment following action observation therapy. They demonstrated a significant change in functional connectivity in the right sensorimotor networks when a significant linguistic improvement was present, suggesting that this therapy improves naming abilities in aphasic patients. Even more recently, Durand et al. (2018), explicitly attributed their rehabilitation approach (Personalized Observation, Execution, and Mental imagery therapy, POEM) to the recent evidence of the embodied framework and identified the neural substrate of their approach via neuroimaging before and after intervention. They combined the potential of action observation, gesture execution and mental imagery into the therapy of two aphasic patients (i.e., *proof of concept* study). Taking into account the preliminary nature of this study, the results showed a positive behavioral outcome for both trained and untrained items, and the neural changes were consistent with an account based on the interaction between the motor and the language systems. The potential of this kind of therapies is promising, yet requires further investigation including control interventions and relevant conditions to better identify the underlying mechanisms both in L1 and L2.

The Semantic Feature Analysis therapy (SFA, Boyle and Coelho, 1995), could also be considered as an experimental manipulation of the motor system, and may also be used in bilingual patients. Similarly to the Action Observation Therapy, the SFA therapy focuses on increasing the activation of semantic features (e.g., action, use, properties) associated with the target word to be retrieved. This intervention has shown a positive correlation between responsiveness to the therapy and the activation of the left precentral gyrus and the left inferior parietal lobule (Marcotte et al., 2012). The left inferior parietal lobule is a multimodal associative area, receiving auditory, visual and somatosensory input (Caspers et al., 2013), and connected to Wernicke's and Broca's areas via the arcuate fasciculus, a white matter tract passing through the precentral gyrus. Based on this, Durand and Ansaldo (2013) took the results from Marcotte et al. (2012) one step further and claimed this path to be recruited during Semantic Feature Analysis therapy, which can in turn lead to positive language production outcomes. For a recent review on the characteristics and effectiveness of SFA therapy results, see Efstratiadou et al. (2018). In terms of bilingualism, Knoph et al. (2015, 2017) were the only ones to measure the effect of SFA therapy in late acquired languages. The authors showed that an overall improvement in verb and narrative production in the treated language could be generalized to the untreated ones in multilingual speakers.

Finally, in regard to the issues mentioned so far, one does wonder whether experimental manipulations of the language system may also produce promising effects on the impaired motor system in monolingual and bilingual patients. Some studies do hint that this may be a promising line of research

(e.g., Maitra et al., 2006). In Maitra et al. (2006), for example, patients that had suffered a stroke had their movements facilitated with self-speech (i.e., self-vocalization). As Anderlini et al. (2019) suggest, the choice of the type of therapeutic approach should consider both the language and motor systems and how they interact, especially when motor and language impairments coexist.

Of course, studies on L2 acquisition may be of special interest in future work on this topic too, as rehabilitation and learning may be grounded on similar mechanisms (e.g., motor areas response to learning the meaning of novel action words in Kiefer et al., 2007; Liuzzi et al., 2010; James and Swain, 2011; Fargier et al., 2012; Bechtold et al., 2018). Still, in sum, embodiment-based therapies offer interesting solutions in L1 and, given the data presented in this review, which assume language-motor association in both L1 and L2, potentially also in L2. In fact, bilingual rehabilitation, the cross language transfer (CLT) of treatment benefits from one language to the other(s) is a notable topic. It is not yet clear which factors influence the success of CLT in bilingual aphasics: premorbid language proficiency, degree and type of language impairments or various forms of therapy (Miertsch et al., 2009; Faroqi-Shah et al., 2010; Kiran and Iakupova, 2011; Kiran et al., 2013; Ansaldo and Saidi, 2014; Radman et al., 2016). Moreover, if the transfer does not take place, the selective recovery of one language could be seen as partial evidence of a different neural representation of the two languages. This issue though, has not yet been explored in the context of embodiment therapies. The engagement of (usually spared) motor areas and the knowledge about the degree of L1 and L2 embodiment could offer new hypotheses about CLT.

THE FUTURE OF L2 EMBODIMENT STUDIES

Theoretical Research

There are many challenging paths in this topic ahead of us, and for any rigorous attempt to better understand lexico-semantic embodiment in L2, we would suggest three critical issues to seriously consider. First, although all studies on the topic have concentrated on a *language-to-motor* directional effect, targeting a *motor-to-language* effect might improve our understanding of the language-motor interaction. This could be addressed by directly changing the excitability of the motor cortex with the application of non-invasive brain stimulation techniques and examining its impact on second language processing. The same goal can be addressed with lesion studies including bilingual patients with motor impairment or including elderly people. As sensory-motor and cognitive functions decline in aging (Baltes and Lindenberger, 1997), the reciprocal influence of these functions could be addressed in monolingual (Vallet, 2015) and bilingual elderly populations. Second, within-participant designs should be favored over between-participant ones. This is crucial in order to minimize the impact of inter-individual sociolinguistic differences, which have been shown to interact with language representations (e.g., De Groot, 1995). Third and

finally – and closely related to the issue of processing stage discussed earlier – measurements and tasks enabling us to specify both space and time characteristics of the mechanisms under investigation should be carefully chosen. For example, functional neuroimaging tools, may provide us with both strength and timing (i.e., onset and duration) of any sensorimotor activation, given they are used in conjunction with the appropriate tasks. More specifically, these tasks should enable us to appropriately access both shallow and deep processing (e.g., lexical and semantic access).

Clinical Research

We believe that this shift in treatment approaches – merging traditional speech and language therapies with a motor integrated perspective – opens new directions in bilingual aphasia rehabilitation. We argue, though, that three necessary issues need to be further addressed and clarified in future studies. First, due to the scarce literature on the subject, additional pre-registered and randomized controlled studies need to be conducted to confirm that therapies based on sensorimotor activation do indeed improve L1 language processing, specifically for sensorimotor-related stimuli in aphasics. Second, clear evidence needs to be provided to show that the same therapy can improve L2 language processing, again, specifically for sensorimotor-related stimuli in aphasics. To our knowledge, only Knoph et al. (2015, 2017) have provided SFA therapy in L2, providing some evidence of improvement in L2. Third and finally, given additional evidence corroborating Knoph and colleagues' findings, therapy outcomes in L2 and L1 would need to be compared and contrasted. Typically, a crossover randomized control trial study could be conducted to address this, provided that the factors influencing therapy outcomes in L1 and L2 (e.g., language competence) are taken into account. Theoretically, it will also bring further enlightenment on differences of the degree of L2 embodiment compared to L1. Clinically, it will bring evidence-based driven awareness in the choice of the therapeutic approach and the language of the therapy. Given that these three issues are rigorously addressed, it should enable us to directly focus on the CLT of therapy outcomes. More specifically, the direction (i.e., L1 to L2, L2 to L1, or both) and magnitude of the transfer could provide us with new insights into the mechanisms underlying embodiment effects. Importantly, we argue that *embodied* therapies could well complement conventional ones – not supplant them –, both still

needing more data for clinicians to choose and apply evidence-based interventions.

CONCLUSION

In light of the exponential increase in multilingual populations worldwide, a better understanding of the mechanisms underlying the interplay between neural structures involved in the processing of more than one language is central. The sensorimotor embodiment account offers an opportunity to further our knowledge in several areas of research, including semantic processing in mono- and bilinguals, language learning, neural mechanisms of language processing and rehabilitation in L2. Overall, all the reviewed studies investigating sensorimotor involvement in semantic processing showed that L2 is – at least to some extent – embodied. Further investigating the factors influencing the degree of L2 embodiment is relevant from a theoretical point of view, of course, but also to confirm or dismiss the value of language therapeutic approaches based on embodiment theories as a complement of speech and language therapies in bilinguals. We have outlined several important issues to tackle in the future, and hope that these will be taken as a sign to encourage rigorous and innovative research in this topic, both in a theoretical and applied perspective.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

This study was supported by the Swiss National Science Foundation (SNF grants 325130_156937/2 and 325130_182594).

ACKNOWLEDGMENTS

We are very grateful to PD Dr. Lucas Spierer for the helpful comments and discussions throughout the preparation of the manuscript. We thank the two reviewers for their valuable input.

REFERENCES

- Abutalebi, J. (2008). Neural aspects of second language representation and language control. *Acta Psychol.* 128, 466–478. doi: 10.1016/j.actpsy.2008.03.014
- Abutalebi, J., Canini, M., Della Rosa, P. A., Green, D. W., and Weekes, B. S. (2015). The neuroprotective effects of bilingualism upon the inferior parietal lobule: a structural neuroimaging study in aging Chinese bilinguals. *J. Neurolinguistics* 33, 3–13. doi: 10.1016/j.jneuroling.2014.09.008
- Adams, A. M. (2016). How language is embodied in bilinguals and children with specific language impairment. *Front. Psychol.* 7:1209. doi: 10.3389/fpsyg.2016.01209
- Aglioti, S. (1999). "Language and memory systems," in *Concise Encyclopedia of Language Pathology*, ed. F. Fabbro (Amsterdam: Elsevier), 371–377.
- Ahlberg, D. K., Bischoff, H., Kaup, B., Bryant, D., and Stroyk, J. V. (2017). Grounded cognition: comparing Language × Space interactions in first language and second language. *Appl. Psycholinguist.* 39, 437–459. doi: 10.1017/S014271641700042X
- Alemanno, F., Houdayer, E., Cursi, M., Velikova, S., Tettamanti, M., Comi, G., et al. (2012). Action-related semantic content and negation polarity modulate motor areas during sentence reading: an event-related desynchronization study. *Brain Res.* 1484, 39–49. doi: 10.1016/j.brainres.2012.09.030

- Anderlini, D., Wallis, G., and Marinovic, W. (2019). Language as a predictor of motor recovery: the case for a more global approach to stroke rehabilitation. *Neurorehabil. Neural Repair* 33, 167–178. doi: 10.1177/1545968319829454
- Anderson, M. L. (2007a). Evolution of cognitive function via redeployment of brain areas. *Neuroscientist* 13, 13–21. doi: 10.1177/1073858406294706
- Anderson, M. L. (2007b). The massive redeployment hypothesis and the functional topography of the brain. *Philos. Psychol.* 20, 143–174. doi: 10.1080/09515080701197163
- Anderson, M. L. (2010). Neural reuse: a fundamental organizational principle of the brain. *Behav. Brain Sci.* 33, 245–266. doi: 10.1017/S0140525X10000853
- Ansaldi, A. I., and Saidi, L. G. (2014). Aphasia therapy in the age of globalization: cross-linguistic therapy effects in bilingual aphasia. *Behav. Neurol.* 2014:603085. doi: 10.1155/2014/603085
- Arévalo, A., Perani, D., Cappa, S. F., Butler, A., Bates, E., and Dronkers, N. (2007). Action and object processing in aphasia: from nouns and verbs to the effect of manipulability. *Brain Lang.* 100, 79–94. doi: 10.1016/j.bandl.2006.06.012
- Aziz-Zadeh, L., Wilson, S. M., Rizzolatti, G., and Iacoboni, M. (2006). Congruent embodied representations for visually presented actions and linguistic phrases describing actions. *Curr. Biol.* 16, 1818–1823. doi: 10.1016/j.cub.2006.07.060
- Bak, T. H. (2013). The neuroscience of action semantics in neurodegenerative brain diseases. *Curr. Opin. Neurol.* 26, 671–677. doi: 10.1097/WCO.0000000000000039
- Bak, T. H., Yancopoulou, D., Nestor, P. J., Xuereb, J. H., Spillantini, M. G., Pulvermüller, F., et al. (2006). Clinical, imaging and pathological correlates of a hereditary deficit in verb and action processing. *Brain* 129, 321–332. doi: 10.1093/brain/awh701
- Baltes, P. B., and Lindenberger, U. (1997). Emergence of a powerful connection between sensory and cognitive functions across the adult life span: a new window to the study of cognitive aging? *Psychol. Aging* 12, 12–21.
- Barsalou, L. W. (2008). Grounded cognition. *Annu. Rev. Psychol.* 59, 617–645. doi: 10.1146/annurev.psych.59.103006.093639
- Barsalou, L. W. (2016). On staying grounded and avoiding quixotic dead ends. *Psychon. Bull. Rev.* 23, 1122–1142. doi: 10.3758/s13423-016-1028-3
- Baumeister, J. C., Foroni, F., Conrad, M., Rumiati, R. I., and Winkielman, P. (2017). Embodiment and emotional memory in first vs. second language. *Front. Psychol.* 8:394. doi: 10.3389/fpsyg.2017.00394
- Bechtold, L., Ghio, M., Lange, J., and Bellebaum, C. (2018). Event-related desynchronization of mu and beta oscillations during the processing of novel tool names. *Brain Lang.* 177–178, 44–55. doi: 10.1016/j.bandl.2018.01.004
- Beilock, S. L., Lyons, I. M., Mattarella-Micke, A., Nusbaum, H. C., and Small, S. L. (2008). Sports experience changes the neural processing of action language. *Proc. Natl. Acad. Sci. U.S.A.* 105, 13269–13273. doi: 10.1073/pnas.0803424105
- Bergen, B., Lau, T.-T. C., Narayan, S., Stojanovic, D., and Wheeler, K. (2010). Body part representations in verbal semantics. *Mem. Cogn.* 38, 969–981. doi: 10.3758/MC.38.7.969
- Bidet-Ildi, C., Meugnot, A., Beauprez, S.-A., Gimenes, M., and Toussaint, L. (2017). Short-term upper limb immobilization affects action-word understanding. *J. Exp. Psychol. Learn. Mem. Cogn.* 43, 1129–1139. doi: 10.1037/xlm0000373
- Birba, A., García-Cordero, I., Kozono, G., Legaz, A., Ibáñez, A., Sedeño, L., et al. (2017). Losing ground: frontostriatal atrophy disrupts language embodiment in Parkinson's and Huntington's disease. *Neurosci. Biobehav. Rev.* 80, 673–687. doi: 10.1016/j.neubiorev.2017.07.011
- Bocanegra, Y., García, A. M., Lopera, F., Pineda, D., Baena, A., Ospina, P., et al. (2017). Unspeakable motion: selective action-verb impairments in Parkinson's disease patients without mild cognitive impairment. *Brain Lang.* 168(Suppl. C), 37–46. doi: 10.1016/j.bandl.2017.01.005
- Bocanegra, Y., García, A. M., Pineda, D., Buritica, O., Villegas, A., Lopera, F., et al. (2015). Syntax, action verbs, action semantics, and object semantics in Parkinson's disease: dissociability, progression, and executive influences. *Cortex* 69(Suppl. C), 237–254. doi: 10.1016/j.cortex.2015.05.022
- Bonifazi, S., Tomaiuolo, F., Altoè, G., Ceravolo, M. G., Provinciali, L., and Marangolo, P. (2013). Action observation as a useful approach for enhancing recovery of verb production: new evidence from aphasia. *Eur. J. Phys. Rehabil. Med.* 49, 473–481.
- Borghì, A. M., and Cimatti, F. (2009). “Words as tools and the problem of abstract word meanings,” in *Proceedings of the 31st Annual Conference of the Cognitive Science Society*, eds N. Taatgen and H. van Rijn (Amsterdam: Cognitive Science Society), 2304–2309.
- Boulenger, V., Hauk, O., and Pulvermüller, F. (2009). Grasping ideas with the motor system: semantic somatotopy in idiom comprehension. *Cereb. Cortex* 19, 1905–1914. doi: 10.1093/cercor/bhn217
- Boulenger, V., Mechtouff, L., Thobois, S., Broussolle, E., Jeannerod, M., and Nazir, T. A. (2008). Word processing in Parkinson's disease is impaired for action verbs but not for concrete nouns. *Neuropsychologia* 46, 743–756. doi: 10.1016/j.neuropsychologia.2007.10.007
- Boyle, M., and Coelho, C. A. (1995). Application of semantic feature analysis as a treatment for aphasic Dysnomia. *Am. J. Speech Lang. Pathol.* 4, 94–98. doi: 10.1044/1058-0360.0404.94
- Branscheidt, M., Hoppe, J., Freundlieb, N., Zwitserlood, P., and Liuzzi, G. (2017a). tDCS over the motor cortex shows differential effects on action and object words in associative word learning in healthy aging. *Front. Aging Neurosci.* 9:137. doi: 10.3389/fnagi.2017.00137
- Branscheidt, M., Hoppe, J., Zwitserlood, P., and Liuzzi, G. (2017b). tDCS over the motor cortex improves lexical retrieval of action words in post-stroke aphasia. *J. Neurophysiol.* 119, 621–630. doi: 10.1152/jn.00285.2017
- Buccino, G., Dalla Volta, R., Arabia, G., Morelli, M., Chiriaco, C., Lupo, A., et al. (2017a). Processing graspable object images and their nouns is impaired in Parkinson's disease patients. *Cortex* 100, 32–39. doi: 10.1016/j.cortex.2017.03.009
- Buccino, G., Marino, B. F., Bulgarelli, C., and Mezzadri, M. (2017b). Fluent speakers of a second language process graspable nouns expressed in L2 like in their native language. *Front. Psychol.* 8:1306. doi: 10.3389/fpsyg.2017.01306
- Buccino, G., and Mezzadri, M. (2015). Embodied language and the process of language learning and teaching. *Emot. Lang.* 10, 191–208. doi: 10.1075/ceb.10.10buc
- Buccino, G., Riggio, L., Melli, G., Binkofski, F., Gallese, V., and Rizzolatti, G. (2005). Listening to action-related sentences modulates the activity of the motor system: a combined TMS and behavioral study. *Cogn. Brain Res.* 24, 355–363. doi: 10.1016/j.cogbrainres.2005.02.020
- Butler, Y. G. (2012). “Bilingualism/multilingualism and second-language acquisition,” in *The Handbook of Bilingualism and Multilingualism*, eds T. K. Bhatia and W. C. Ritchie (Hoboken, NJ: John Wiley & Sons, Ltd), 109–136. doi: 10.1002/9781118332382.ch5
- Caldwell-Harris, C. L. (2015). Emotionality differences between a native and foreign language: implications for everyday life. *Curr. Dir. Psychol. Sci.* 24, 214–219. doi: 10.1177/0963721414566268
- Caramazza, A., Anzellotti, S., Strnad, L., and Lingnau, A. (2014). Embodied cognition and mirror neurons: a critical assessment. *Annu. Rev. Neurosci.* 37, 1–15. doi: 10.1146/annurev-neuro-071013-013950
- Cardona, J. F., Gershanik, O., Gelormini-Lezama, C., Houck, A. L., Cardona, S., Kargieman, L., et al. (2013). Action-verb processing in Parkinson's disease: new pathways for motor-language coupling. *Brain Struct. Funct.* 218, 1355–1373. doi: 10.1007/s00429-013-0510-1
- Caspers, S., Schleicher, A., Bacha-Trams, M., Palomero-Gallagher, N., Amunts, K., and Zilles, K. (2013). Organization of the human inferior parietal lobule based on receptor architectonics. *Cereb. Cortex* 23, 615–628. doi: 10.1093/cercor/bhs048
- Cotelli, M., Borroni, B., Manenti, R., Zanetti, M., Arévalo, A., Cappa, S. F., et al. (2007). Action and object naming in Parkinson's disease without dementia. *Eur. J. Neurol.* 14, 632–637. doi: 10.1111/j.1468-1331.2007.01797.x
- Cotelli, M., Manenti, R., Brambilla, M., and Borroni, B. (2018). The role of the motor system in action naming in patients with neurodegenerative extrapyramidal syndromes. *Cortex* 100, 191–214. doi: 10.1016/j.cortex.2017.05.011
- Dahl, A., and Vulchanova, M. D. (2014). Naturalistic acquisition in an early language classroom. *Front. Psychol.* 5:329. doi: 10.3389/fpsyg.2014.00329
- Davis, J. D., Winkielman, P., and Coulson, S. (2015). Facial action and emotional language: ERP evidence that blocking facial feedback selectively impairs sentence comprehension. *J. Cogn. Neurosci.* 27, 2269–2280. doi: 10.1162/jocn_a_00858
- De Grauwe, S., Willems, R. M., Rueschemeyer, S.-A., Lemhöfer, K., and Schriefers, H. (2014). Embodied language in first- and second-language speakers: neural correlates of processing motor verbs. *Neuropsychologia* 56(Suppl. C), 334–349. doi: 10.1016/j.neuropsychologia.2014.02.003

- De Groot, A. M. B. (1995). Determinants of bilingual lexicosemantic organisation. *Comput. Assist. Lang. Learn.* 8, 151–180. doi: 10.1080/0958822940080204
- Degani, T., Prior, A., and Hajjara, W. (2018). Cross-language semantic influences in different script bilinguals. *Biling. Lang. Cogn.* 21, 782–804. doi: 10.1017/S1366728917000311
- Dehaene, S., and Cohen, L. (2007). Cultural recycling of cortical maps. *Neuron* 56, 384–398. doi: 10.1016/j.neuron.2007.10.004
- Dekker, T. M., Mareschal, D., Johnson, M. H., and Sereno, M. I. (2014). Picturing words? Sensorimotor cortex activation for printed words in child and adult readers. *Brain Lang.* 139, 58–67. doi: 10.1016/j.bandl.2014.09.009
- Desai, R. H., Herter, T., Riccardi, N., Rorden, C., and Fridriksson, J. (2015). Concepts within reach: action performance predicts action language processing in stroke. *Neuropsychologia* 71(Suppl. C), 217–224. doi: 10.1016/j.neuropsychologia.2015.04.006
- Difrancesco, S., Pulvermüller, F., and Mohr, B. (2012). Intensive language-action therapy (ILAT): the methods. *Aphasiology* 26, 1317–1351. doi: 10.1080/02687038.2012.705815
- Dijkstra, T., and van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: from identification to decision. *Biling. Lang. Cogn.* 5, 175–197. doi: 10.1017/S1366728902003012
- Dreyer, F. R., and Pulvermüller, F. (2018). Abstract semantics in the motor system? – An event-related fMRI study on passive reading of semantic word categories carrying abstract emotional and mental meaning. *Cortex* 100, 52–70. doi: 10.1016/j.cortex.2017.10.021
- Druks, J., and Weekes, B. S. (2013). Parallel deterioration to language processing in a bilingual speaker. *Cogn. Neuropsychol.* 30, 578–596. doi: 10.1080/02643294.2014.882814
- Dudschig, C., de la Vega, I., and Kaup, B. (2014). Embodiment and second-language: automatic activation of motor responses during processing spatially associated L2 words and emotion L2 words in a vertical Stroop paradigm. *Brain Lang.* 132(Suppl. C), 14–21. doi: 10.1016/j.bandl.2014.02.002
- Durand, E., and Ansaldo, A. I. (2013). Recovery from anomia following Semantic Feature Analysis: therapy-induced neuroplasticity relies upon a circuit involving motor and language processing areas. *Ment. Lex.* 8, 195–215. doi: 10.1075/ml.8.2.04dur
- Durand, E., Berroir, P., and Ansaldo, A. I. (2018). The neural and behavioral correlates of anomia recovery following personalized observation, execution, and mental imagery therapy: a proof of concept. *Neural Plast.* 2018:5943759. doi: 10.1155/2018/5943759
- Efstratiadou, E. A., Papanthasiou, I., Holland, R., Archonti, A., and Hilari, K. (2018). A systematic review of semantic feature analysis therapy studies for aphasia. *J. Speech Lang. Hear. Res.* 61, 1261–1278. doi: 10.1044/2018_JSLHR-16-0330
- Eilola, T. M., and Havelka, J. (2011). Behavioural and physiological responses to the emotional and taboo Stroop tasks in native and non-native speakers of English. *Int. J. Biling.* 15, 353–369. doi: 10.1177/1367006910379263
- Elsner, B., Kugler, J., Pohl, M., and Mehrholz, J. (2013). Transcranial direct current stimulation (tDCS) for improving function and activities of daily living in patients after stroke. *Cochrane Database Syst. Rev.* 11:CD009645. doi: 10.1002/14651858.CD009645.pub2
- Fargier, R., Paulignan, Y., Boulenger, V., Monaghan, P., Reboul, A., and Nazir, T. A. (2012). Learning to associate novel words with motor actions: language-induced motor activity following short training. *Cortex* 48, 888–899. doi: 10.1016/j.cortex.2011.07.003
- Faroqi-Shah, Y., Frymark, T., Mullen, R., and Wang, B. (2010). Effect of treatment for bilingual individuals with aphasia: a systematic review of the evidence. *J. Neurolinguistics* 23, 319–341. doi: 10.1016/j.jneuroling.2010.01.002
- Fernandino, L., Conant, L. L., Binder, J. R., Blindauer, K., Hiner, B., Spangler, K., et al. (2013a). Parkinson's disease disrupts both automatic and controlled processing of action verbs. *Brain Lang.* 127, 65–74. doi: 10.1016/j.bandl.2012.07.008
- Fernandino, L., Conant, L. L., Binder, J. R., Blindauer, K., Hiner, B., Spangler, K., et al. (2013b). Where is the action? Action sentence processing in Parkinson's disease. *Neuropsychologia* 51, 1510–1517. doi: 10.1016/j.neuropsychologia.2013.04.008
- Finkbeiner, M., Forster, K., Nicol, J., and Nakamura, K. (2004). The role of polysemy in masked semantic and translation priming. *J. Mem. Lang.* 51, 1–22. doi: 10.1016/j.jml.2004.01.004
- Fino, E., Menegatti, M., Avenanti, A., and Rubini, M. (2016). Enjoying vs. smiling: facial muscular activation in response to emotional language. *Biol. Psychol.* 118(Suppl. C), 126–135. doi: 10.1016/j.biopsycho.2016.04.069
- Froni, F. (2015). Do we embody second language? Evidence for “partial” simulation during processing of a second language. *Brain Cogn.* 99(Suppl. C), 8–16. doi: 10.1016/j.bandc.2015.06.006
- Froni, F., and Semin, G. R. (2009). Language that puts you in touch with your bodily feelings: the multimodal responsiveness of affective expressions. *Psychol. Sci.* 20, 974–980. doi: 10.1111/j.1467-9280.2009.02400.x
- Froni, F., and Semin, G. R. (2013). Comprehension of action negation involves inhibitory simulation. *Front. Hum. Neurosci.* 7:209. doi: 10.3389/fnhum.2013.00209
- Gallese, V. (2008). Mirror neurons and the social nature of language: the neural exploitation hypothesis. *Soc. Neurosci.* 3, 317–333. doi: 10.1080/17470910701563608
- Gallese, V., and Cuccio, V. (2018). The neural exploitation hypothesis and its implications for an embodied approach to language and cognition: insights from the study of action verbs processing and motor disorders in Parkinson's disease. *Cortex* 100, 215–225. doi: 10.1016/j.cortex.2018.01.010
- Gallese, V., and Lakoff, G. (2005). The Brain's concepts: the role of the Sensory-motor system in conceptual knowledge. *Cogn. Neuropsychol.* 22, 455–479. doi: 10.1080/02643290442000310
- García, A., and Ibáñez, A. (2014). Words in motion: motor-language coupling in Parkinson's disease. *Transl. Neurosci.* 5, 152–159. doi: 10.2478/s13380-014-0218-6
- García, A. M., Bocanegra, Y., Herrera, E., Moreno, L., Carmona, J., Baena, A., et al. (2018). Parkinson's disease compromises the appraisal of action meanings evoked by naturalistic texts. *Cortex* 100, 111–126. doi: 10.1016/j.cortex.2017.07.003
- García, A. M., Carrillo, F., Orozco-Arroyave, J. R., Trujillo, N., Vargas Bonilla, J. F., Fittipaldi, S., et al. (2016). How language flows when movements don't: an automated analysis of spontaneous discourse in Parkinson's disease. *Brain Lang.* 162, 19–28. doi: 10.1016/j.bandl.2016.07.008
- García, A. M., and Ibáñez, A. (2016). A touch with words: dynamic synergies between manual actions and language. *Neurosci. Biobehav. Rev.* 68(Suppl. C), 59–95. doi: 10.1016/j.neubiorev.2016.04.022
- García, A. M., Sedeño, L., Trujillo, N., Bocanegra, Y., Gomez, D., Pineda, D., et al. (2017). Language deficits as a preclinical window into Parkinson's disease: evidence from asymptomatic parkin and dardarin mutation carriers. *J. Int. Neuropsychol. Soc.* 23, 150–158. doi: 10.1017/S1355617716000710
- García-Gómez, A. B., and Macizo, P. (2018). Learning nouns and verbs in a foreign language: the role of gestures. *Appl. Psycholinguist.* 40, 473–507. doi: 10.1017/S0142716418000656
- Ghazi-Saidi, L., and Ansaldo, A. I. (2017). The neural correlates of semantic and phonological transfer effects: language distance matters. *Biling. Lang. Cogn.* 20, 1080–1094. doi: 10.1017/S136672891600064X
- Ghio, M., Locatelli, M., Tettamanti, A., Perani, D., Gatti, R., and Tettamanti, M. (2018). Cognitive training with action-related verbs induces neural plasticity in the action representation system as assessed by gray matter brain morphometry. *Neuropsychologia* 114, 186–194. doi: 10.1016/j.neuropsychologia.2018.04.036
- Ghio, M., and Tettamanti, M. (2016). “Chapter 52 - grounding sentence processing in the sensory-motor system,” in *Neurobiology of Language*, eds G. Hickok and S. L. Small (San Diego, CA: Academic Press), 647–657.
- Gianelli, C., and Dalla Volta, R. (2015). Does listening to action-related sentences modulate the activity of the motor system? Replication of a combined TMS and behavioral study. *Front. Psychol.* 5:1511. doi: 10.3389/fpsyg.2014.01511
- Gijssels, T., Ivry, R. B., and Casasanto, D. (2018). tDCS to premotor cortex changes action verb understanding: complementary effects of inhibitory and excitatory stimulation. *Sci. Rep.* 8:11452. doi: 10.1038/s41598-018-29600-6
- Gili, T., Fiori, V., Pasquale, G. D., Sabatini, U., Caltagirone, C., and Marangolo, P. (2017). Right sensory-motor functional networks subserve action observation therapy in aphasia. *Brain Imaging Behav.* 11, 1397–1411. doi: 10.1007/s11682-016-9635-1
- Glenberg, A. M. (1997). What memory is for. *Behav. Brain Sci.* 20, 1–19. doi: 10.1017/S0140525X97000010
- Glenberg, A. M., and Gallese, V. (2012). Action-based language: a theory of language acquisition, comprehension, and production. *Cortex* 48, 905–922. doi: 10.1016/j.cortex.2011.04.010

- Glenberg, A. M., and Kaschak, M. P. (2002). Grounding language in action. *Psychon. Bull. Rev.* 9, 558–565. doi: 10.3758/BF03196313
- Glenberg, A. M., Sato, M., and Cattaneo, L. (2008). Use-induced motor plasticity affects the processing of abstract and concrete language. *Curr. Biol.* 18, R290–R291. doi: 10.1016/j.cub.2008.02.036
- Goldinger, S. D., Papesch, M. H., Barnhart, A. S., Hansen, W. A., and Hout, M. C. (2016). The poverty of embodied cognition. *Psychon. Bull. Rev.* 23, 959–978. doi: 10.3758/s13423-015-0860-1
- Goldin-Meadow, S., Nusbaum, H., Kelly, S. D., and Wagner, S. (2001). Explaining math: gesturing lightens the load. *Psychol. Sci.* 12, 516–522. doi: 10.1111/1467-9280.00395
- Gough, P. M., Campione, G. C., and Buccino, G. (2013). Fine tuned modulation of the motor system by adjectives expressing positive and negative properties. *Brain Lang.* 125, 54–59. doi: 10.1016/j.bandl.2013.01.012
- Hamrick, P., Lum, J. A. G., and Ullman, M. T. (2018). Child first language and adult second language are both tied to general-purpose learning systems. *Proc. Natl. Acad. Sci. U.S.A.* 115, 1487–1492. doi: 10.1073/pnas.1713975115
- Harris, C. L. (2004). Bilingual speakers in the lab: psychophysiological measures of emotional reactivity. *J. Multilin. Multicult. Dev.* 25, 223–247. doi: 10.1080/01434630408666530
- Harris, C. L., Ayçiçeği, A., and Gleason, J. B. (2003). Taboo words and reprimands elicit greater autonomic reactivity in a first language than in a second language. *Appl. Psycholinguist.* 24, 561–579. doi: 10.1017/S0142716403000286
- Hauk, O., Johnsrude, L., and Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron* 41, 301–307. doi: 10.1016/S0896-6273(03)00838-9
- Havas, D. A., Glenberg, A. M., Gutowski, K. A., Lucarelli, M. J., and Davidson, R. J. (2010). Cosmetic use of botulinum toxin-a affects processing of emotional language. *Psychol. Sci.* 21, 895–900. doi: 10.1177/0956797610374742
- Havas, D. A., Glenberg, A. M., and Rinck, M. (2007). Emotion simulation during language comprehension. *Psychon. Bull. Rev.* 14, 436–441. doi: 10.3758/BF03194085
- Hernandez, A. E., and Li, P. (2007). Age of acquisition: its neural and computational mechanisms. *Psychol. Bull.* 133, 638–650. doi: 10.1037/0033-2909.133.4.638
- Hill, E. L. (2001). Non-specific nature of specific language impairment: a review of the literature with regard to concomitant motor impairments. *Int. J. Lang. Commun. Disord.* 36, 149–171. doi: 10.1080/13682820010019874
- Hsu, C.-T., Jacobs, A. M., and Conrad, M. (2015). Can Harry Potter still put a spell on us in a second language? An fMRI study on reading emotion-laden literature in late bilinguals. *Cortex* 63(Suppl. C), 282–295. doi: 10.1016/j.cortex.2014.09.002
- Hurley, S. (2008). The shared circuits model (SCM): how control, mirroring, and simulation can enable imitation, deliberation, and mindreading. *Behav. Brain Sci.* 31, 1–22; discussion 22–58. doi: 10.1017/S0140525X07003123
- Hurley, S. L., and Chater, N. (2005). *Perspectives on Imitation: Mechanisms of Imitation and Imitation in Animals*. Cambridge, MA: MIT Press.
- Inkster, M., Wellsby, M., Lloyd, E., and Pexman, P. M. (2016). Development of embodied word meanings: sensorimotor effects in children's lexical processing. *Front. Psychol.* 7:317. doi: 10.3389/fpsyg.2016.00317
- Innocenti, A., Stefani, E. D., Sestito, M., and Gentilucci, M. (2014). Understanding of action-related and abstract verbs in comparison: a behavioral and TMS study. *Cogn. Process.* 15, 85–92. doi: 10.1007/s10339-013-0583-z
- Isel, F., Baumgaertner, A., Thrän, J., Meisel, J. M., and Büchel, C. (2010). Neural circuitry of the bilingual mental lexicon: effect of age of second language acquisition. *Brain Cogn.* 72, 169–180. doi: 10.1016/j.bandc.2009.07.008
- Izura, C., and Ellis, A. W. (2004). Age of acquisition effects in translation judgement tasks. *J. Mem. Lang.* 50, 165–181. doi: 10.1016/j.jml.2003.09.004
- James, K. H., and Swain, S. N. (2011). Only self-generated actions create sensorimotor systems in the developing brain. *Dev. Sci.* 14, 673–678. doi: 10.1111/j.1467-7687.2010.01011.x
- Jared, D., Poh, R. P. Y., and Paivio, A. (2013). L1 and L2 picture naming in Mandarin-English bilinguals: a test of bilingual dual coding theory. *Biling. Lang. Cogn.* 16, 383–396. doi: 10.1017/S1366728912000685
- Jirak, D., Menz, M. M., Buccino, G., Borghi, A. M., and Binkofski, F. (2010). Grasping language – A short story on embodiment. *Conscious. Cogn.* 19, 711–720. doi: 10.1016/j.concog.2010.06.020
- Johari, K., Ashrafi, F., Zali, A., Ashayeri, H., Fabbro, F., and Zanini, S. (2013). Grammatical deficits in bilingual Azari – Farsi patients with Parkinson's disease. *J. Neurolinguistics* 26, 22–30. doi: 10.1016/j.jneuroling.2012.02.004
- Johnson, J. S., and Newport, E. L. (1989). Critical period effects in second language learning: the influence of maturational state on the acquisition of English as a second language. *Cogn. Psychol.* 21, 60–99. doi: 10.1016/0010-0285(89)90003-0
- Kargieman, L., Herrera, E., Baez, S., García, A. M., Dottori, M., Gelormini, C., et al. (2014). Motor–language coupling in Huntington's disease families. *Front. Aging Neurosci.* 6:122. doi: 10.3389/fnagi.2014.00122
- Kemmerer, D., Rudrauf, D., Manzel, K., and Tranel, D. (2012). Behavioral patterns and lesion sites associated with impaired processing of lexical and conceptual knowledge of actions. *Cortex* 48, 826–848. doi: 10.1016/j.cortex.2010.11.001
- Khateb, A., Pegna, A. J., Michel, C. M., Mouthon, M., and Annoni, J.-M. (2016). Semantic relatedness and first-second language effects in the bilingual brain: a brain mapping study. *Biling. Lang. Cogn.* 19, 311–330. doi: 10.1017/S1366728915000140
- Kiefer, M., Sim, E.-J., Liebich, S., Hauk, O., and Tanaka, J. (2007). Experience-dependent plasticity of conceptual representations in human sensory-motor areas. *J. Cogn. Neurosci.* 19, 525–542. doi: 10.1162/jocn.2007.19.3.525
- Kiran, S., and Iakupova, R. (2011). Understanding the relationship between language proficiency, language impairment and rehabilitation: evidence from a case study. *Clin. Linguist. Phon.* 25, 565–583. doi: 10.3109/02699206.2011.566664
- Kiran, S., Sandberg, C., Gray, T., Ascenso, E., and Kester, E. (2013). Rehabilitation in bilingual aphasia: evidence for within- and between-language generalization. *Am. J. Speech Lang. Pathol.* 22, S298–S309. doi: 10.1044/1058-0360(2013)12-0085
- Knop, M. I. N., Lind, M., and Simonsen, H. G. (2015). Semantic feature analysis targeting verbs in a quadrilingual speaker with aphasia. *Aphasiology* 29, 1473–1496. doi: 10.1080/02687038.2015.1049583
- Knop, M. I. N., Simonsen, H. G., and Lind, M. (2017). Cross-linguistic transfer effects of verb-production therapy in two cases of multilingual aphasia. *Aphasiology* 31, 1482–1509. doi: 10.1080/02687038.2017.1358447
- Krauss, R. M. (1998). Why do we gesture when we speak? *Curr. Dir. Psychol. Sci.* 7, 54–60. doi: 10.1111/1467-8721.ep13175642
- Kroll, J. F., and Stewart, E. (1994). Category interference in translation and picture naming: evidence for asymmetric connections between bilingual memory representations. *J. Mem. Lang.* 33, 149–174. doi: 10.1016/j.jecp.2008.10.004
- Laganaro, M., and Overton Venet, M. (2001). Acquired alexia in multilingual aphasia and computer-assisted treatment in both languages: issues of generalisation and transfer. *Folia Phoniatri. Logop.* 53, 135–144. doi: 10.1159/000052668
- Laufer, B., and Nation, P. (1999). A vocabulary-size test of controlled productive ability. *Lang. Test.* 16, 33–51. doi: 10.1177/026553229901600103
- Lefaucheur, J.-P., Antal, A., Ayache, S. S., Benninger, D. H., Brunelin, J., Cogiamanian, F., et al. (2017). Evidence-based guidelines on the therapeutic use of transcranial direct current stimulation (tDCS). *Clin. Neurophysiol.* 128, 56–92. doi: 10.1016/j.clinph.2016.10.087
- Lemhöfer, K., and Broersma, M. (2012). Introducing LexTALE: a quick and valid lexical test for advanced learners of english. *Behav. Res. Methods* 44, 325–343. doi: 10.3758/s13428-011-0146-0
- Leonard, M. K., Brown, T. T., Travis, K. E., Gharapetian, L., Hagler, D. J., Dale, A. M., et al. (2010). Spatiotemporal dynamics of bilingual word processing. *Neuroimage* 49, 3286–3294. doi: 10.1016/j.neuroimage.2009.12.009
- Linck, J. A., Kroll, J. F., and Sunderman, G. (2009). Losing access to the native language while immersed in a second language evidence for the role of inhibition in second-language learning. *Psychol. Sci.* 20, 1507–1515. doi: 10.1111/j.1467-9280.2009.02480.x
- Lindgren, E., and Muñoz, C. (2013). The influence of exposure, parents, and linguistic distance on young European learners' foreign language comprehension. *Int. J. Multiling.* 10, 105–129. doi: 10.1080/14790718.2012.679275
- Liuzzi, G., Freundlieb, N., Ridder, V., Hoppe, J., Heise, K., Zimmerman, M., et al. (2010). The involvement of the left motor cortex in learning of a novel action word lexicon. *Curr. Biol.* 20, 1745–1751. doi: 10.1016/j.cub.2010.08.034
- Llama, R., Cardoso, W., and Collins, L. (2010). The influence of language distance and language status on the acquisition of L3 phonology. *Int. J. Multiling.* 7, 39–57. doi: 10.1080/14790710902972255

- Locatelli, M., Gatti, R., and Tettamanti, M. (2012). Training of manual actions improves language understanding of semantically related action sentences. *Front. Psychol.* 3:547. doi: 10.3389/fpsyg.2012.00547
- Macedonia, M. (2014). Bringing back the body into the mind: gestures enhance word learning in foreign language. *Front. Psychol.* 5:1467. doi: 10.3389/fpsyg.2014.01467
- Macedonia, M., and Mueller, K. (2016). Exploring the neural representation of novel words learned through enactment in a word recognition task. *Front. Psychol.* 7:953. doi: 10.3389/fpsyg.2016.00953
- Mahon, B. Z., and Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *J. Physiol. Paris* 102, 59–70. doi: 10.1016/j.jphysparis.2008.03.004
- Maieron, M., Marin, D., Fabbro, F., and Skrap, M. (2013). Seeking a bridge between language and motor cortices: a PPI study. *Front. Hum. Neurosci.* 7:249. doi: 10.3389/fnhum.2013.00249
- Maitra, K. K., Telage, K. M., and Rice, M. S. (2006). Self-speech-induced facilitation of simple reaching movements in persons with stroke. *Am. J. Occup. Ther.* 60, 146–154. doi: 10.5014/ajot.60.2.146
- Marangolo, P., Bonifazi, S., Tomaiuolo, F., Craighero, L., Coccia, M., Altoè, G., et al. (2010). Improving language without words: first evidence from aphasia. *Neuropsychologia* 48, 3824–3833. doi: 10.1016/j.neuropsychologia.2010.09.025
- Marangolo, P., Fiori, V., Sabatini, U., De Pasquale, G., Razzano, C., Caltagirone, C., et al. (2016). Bilateral transcranial direct current stimulation language treatment enhances functional connectivity in the left hemisphere: preliminary data from aphasia. *J. Cogn. Neurosci.* 28, 724–738. doi: 10.1162/jocn_a_00927
- Marcotte, K., Adrover-Roig, D., Damien, B., de Préaumont, M., Gèneux, S., Hubert, M., et al. (2012). Therapy-induced neuroplasticity in chronic aphasia. *Neuropsychologia* 50, 1776–1786. doi: 10.1016/j.neuropsychologia.2012.04.001
- Marian, V., Blumenfeld, H. K., and Kaushanskaya, M. (2007). The language experience and proficiency questionnaire (LEAP-Q): assessing language profiles in bilinguals and multilinguals. *J. Speech Lang. Hear. Res.* 50, 940–967. doi: 10.1044/1092-4388(2007)067
- Marino, B. F. M., Sirianni, M., Volta, R. D., Magliocco, F., Silipo, F., Quattrone, A., et al. (2014). Viewing photos and reading nouns of natural graspable objects similarly modulate motor responses. *Front. Hum. Neurosci.* 8:968. doi: 10.3389/fnhum.2014.00968
- Martin, A. (2016). GRAPES—Grounding representations in action, perception, and emotion systems: how object properties and categories are represented in the human brain. *Psychon. Bull. Rev.* 23, 979–990. doi: 10.3758/s13423-015-0842-3
- Meinzer, M., Darkow, R., Lindenberg, R., and Flöel, A. (2016). Electrical stimulation of the motor cortex enhances treatment outcome in post-stroke aphasia. *Brain* 139, 1152–1163. doi: 10.1093/brain/aww002
- Meteyard, L., Cuadrado, S. R., Bahrami, B., and Vigliocco, G. (2012). Coming of age: a review of embodiment and the neuroscience of semantics. *Cortex* 48, 788–804. doi: 10.1016/j.cortex.2010.11.002
- Miertsch, B., Meisel, J. M., and Isel, F. (2009). Non-treated languages in aphasia therapy: of polyglots benefit from improvement in the treated language. *J. Neurolinguistics* 2, 135–150. doi: 10.1016/j.jneuroling.2008.07.003
- Moon, J., and Jiang, N. (2012). Non-selective lexical access in different-script bilinguals. *Biling. Lang. Cogn.* 15, 173–180. doi: 10.1017/S1366728911000022
- Moreno, E. M., and Kutas, M. (2005). Processing semantic anomalies in two languages: an electrophysiological exploration in both languages of Spanish-English bilinguals. *Cogn. Brain Res.* 22, 205–220. doi: 10.1016/j.cogbrainres.2004.08.010
- Newman, A. J., Tremblay, A., Nichols, E. S., Neville, H. J., and Ullman, M. T. (2011). The influence of language proficiency on lexical semantic processing in native and late learners of English. *J. Cogn. Neurosci.* 24, 1205–1223. doi: 10.1162/jocn_a_00143
- Öttl, B., Dudschig, C., and Kaup, B. (2017). Forming associations between language and sensorimotor traces during novel word learning. *Lang. Cogn.* 9, 156–171. doi: 10.1017/langcog.2016.5
- Papeo, L., Corradi-Dell'Acqua, C., and Rumiati, R. I. (2011). “She” is not like “I”: the tie between language and action is in our imagination. *J. Cogn. Neurosci.* 23, 3939–3948. doi: 10.1162/jocn_a_00075
- Papeo, L., Negri, G. A. L., Zadini, A., and Rumiati, R. I. (2010). Action performance and action-word understanding: evidence of double dissociations in left-damaged patients. *Cogn. Neuropsychol.* 27, 428–461. doi: 10.1080/02643294.2011.570326
- Papeo, L., Pascual-Leone, A., and Caramazza, A. (2013). Disrupting the brain to validate hypotheses on the neurobiology of language. *Front. Hum. Neurosci.* 7:148. doi: 10.3389/fnhum.2013.00148
- Papeo, L., Vallesi, A., Isaja, A., and Rumiati, R. I. (2009). Effects of TMS on different stages of motor and non-motor verb processing in the primary motor cortex. *PLoS One* 4:e4508. doi: 10.1371/journal.pone.0004508
- Paradis, M. (2004). *A Neurolinguistic Theory of Bilingualism*. Amsterdam: John Benjamins Publishing. doi: 10.1075/sibil.18
- Pavlenko, A. (2007). *Emotions and Multilingualism*. Cambridge: Cambridge University Press.
- Pavlenko, A. (2012). Affective processing in bilingual speakers: disembodied cognition? *Int. J. Psychol.* 47, 405–428. doi: 10.1080/00207594.2012.743665
- Pavlenko, A. (2017). Do you wish to waive your rights? Affect and decision-making in multilingual speakers. *Curr. Opin. Psychol.* 17(Suppl. C), 74–78. doi: 10.1016/j.copsyc.2017.06.005
- Perani, D., and Abutalebi, J. (2005). The neural basis of first and second language processing. *Curr. Opin. Neurobiol.* 15, 202–206. doi: 10.1016/j.conb.2005.03.007
- Perani, D., Abutalebi, J., Paulesu, E., Brambati, S., Scifo, P., Cappa, S. F., et al. (2003). The role of age of acquisition and language usage in early, high-proficient bilinguals: an fMRI study during verbal fluency. *Hum. Brain Mapp.* 19, 170–182. doi: 10.1002/hbm.10110
- Persici, V., Vihman, M., Burro, R., and Majorano, M. (2019). Lexical access and competition in bilingual children: the role of proficiency and the lexical similarity of the two languages. *J. Exp. Child Psychol.* 179, 103–125. doi: 10.1016/j.jecp.2018.10.002
- Postle, N., Ashton, R., McFarland, K., and de Zubicaray, G. I. (2013). No specific role for the manual motor system in processing the meanings of words related to the hand. *Front. Hum. Neurosci.* 7:11. doi: 10.3389/fnhum.2013.00011
- Pulvermüller, F., Hauk, O., Nikulin, V. V., and Ilmoniemi, R. J. (2005). Functional links between motor and language systems. *Eur. J. Neurosci.* 21, 793–797. doi: 10.1111/j.1460-9568.2005.03900.x
- Pulvermüller, F., Neininger, B., Elbert, T., Mohr, B., Rockstroh, B., Koebbel, P., et al. (2001). Constraint-induced therapy of chronic aphasia after stroke. *Stroke* 32, 1621–1626. doi: 10.1161/01.STR.32.7.1621
- Qian, W. (2016). Embodied cognition processing and representation of power words by second language learners with different proficiency levels. *Chin. J. Appl. Linguist.* 39, 484–494. doi: 10.1515/cjal-2016-0030
- Radman, N., Spierer, L., Laganaro, M., Annoni, J.-M., and Colombo, F. (2016). Language specificity of lexical-phonological therapy in bilingual aphasia: a clinical and electrophysiological study. *Neuropsychol. Rehabil.* 26, 532–557. doi: 10.1080/09602011.2015.1047383
- Repetto, C., Pedrol, E., and Macedonia, M. (2017). Enrichment effects of gestures and pictures on abstract words in a second language. *Front. Psychol.* 8:2136. doi: 10.3389/fpsyg.2017.02136
- Rose, M. L. (2006). The utility of arm and hand gestures in the treatment of aphasia. *Adv. Speech Lang. Pathol.* 8, 92–109. doi: 10.1080/14417040600657948
- Rose, M. L., Raymer, A. M., Lanyon, L. E., and Attard, M. C. (2013). A systematic review of gesture treatments for post-stroke aphasia. *Aphasiology* 27, 1090–1127. doi: 10.1080/02687038.2013.805726
- Routhier, S., Bier, N., and Macoir, J. (2015). The contrast between cueing and/or observation in therapy for verb retrieval in post-stroke aphasia. *J. Commun. Disord.* 54, 43–55. doi: 10.1016/j.jcomdis.2015.01.003
- Sabourin, L., Brien, C., and Burkholder, M. (2014). The effect of age of L2 acquisition on the organization of the bilingual lexicon: evidence from masked priming. *Biling. Lang. Cogn.* 17, 542–555. doi: 10.1017/S1366728913000643
- Sanjeevan, T., Rosenbaum, D. A., Miller, C., van Hell, J. G., Weiss, D. J., and Mainela-Arnold, E. (2015). Motor issues in specific language impairment: a window into the underlying impairment. *Curr. Dev. Disord. Rep.* 2, 228–236. doi: 10.1007/s40474-015-0051-9
- Scorilli, C., Binkofski, F., Buccino, G., Nicoletti, R., Riggio, L., and Borghi, A. M. (2011). Abstract and concrete sentences, embodiment, and languages. *Front. Psychol.* 2:227. doi: 10.3389/fpsyg.2011.00227
- Sell, A. J., and Kaschak, M. P. (2011). Processing time shifts affects the execution of motor responses. *Brain Lang.* 117, 39–44. doi: 10.1016/j.bandl.2010.07.003
- Sheikh, N. A., and Titone, D. (2016). The embodiment of emotional words in a second language: an eye-movement study. *Cogn. Emot.* 30, 488–500. doi: 10.1080/02699931.2015.1018144

- Sheikh, N. A., and Titone, D. A. (2013). Sensorimotor and linguistic information attenuate emotional word processing benefits: an eye-movement study. *Emotion* 13, 1107–1121. doi: 10.1037/a0032417
- Silverberg, S., and Samuel, A. G. (2004). The effect of age of second language acquisition on the representation and processing of second language words. *J. Mem. Lang.* 51, 381–398. doi: 10.1016/j.jml.2004.05.003
- Stahl, B., Mohr, B., Dreyer, F. R., Lucchese, G., and Pulvermüller, F. (2016). Using language for social interaction: communication mechanisms promote recovery from chronic non-fluent aphasia. *Cortex* 85, 90–99. doi: 10.1016/j.cortex.2016.09.021
- Stein, M., Winkler, C., Kaiser, A., and Dierks, T. (2014). Structural brain changes related to bilingualism: does immersion make a difference? *Front. Psychol.* 5:1116. doi: 10.3389/fpsyg.2014.01116
- Swaminathan, S., MacSweeney, M., Boyles, R., Waters, D., Watkins, K. E., and Möttönen, R. (2013). Motor excitability during visual perception of known and unknown spoken languages. *Brain Lang.* 126, 1–7. doi: 10.1016/j.bandl.2013.03.002
- Taylor, L. J., Evans, C., Greer, J., Senior, C., Coventry, K. R., and Ietswaart, M. (2017). Dissociation between semantic representations for motion and action verbs: evidence from patients with left hemisphere lesions. *Front. Hum. Neurosci.* 11:35. doi: 10.3389/fnhum.2017.00035
- Tettamanti, M., Buccino, G., Saccuman, M. C., Gallese, V., Danna, M., Scifo, P., et al. (2005). Listening to action-related sentences activates fronto-parietal motor circuits. *J. Cogn. Neurosci.* 17, 273–281. doi: 10.1162/0898929053124965
- Tremblay, P., Sato, M., and Small, S. L. (2012). TMS-induced modulation of action sentence priming in the ventral premotor cortex. *Neuropsychologia* 50, 319–326. doi: 10.1016/j.neuropsychologia.2011.12.002
- Trevisan, P., Sedeño, L., Birba, A., Ibáñez, A., and García, A. M. (2017). A moving story: whole-body motor training selectively improves the appraisal of action meanings in naturalistic narratives. *Sci. Rep.* 7:12538. doi: 10.1038/s41598-017-12928-w
- Ullman, M. T. (2001). A neurocognitive perspective on language: the declarative/procedural model. *Nat. Rev. Neurosci.* 2, 717–726. doi: 10.1038/35094573
- Ullman, M. T. (2004). Contributions of memory circuits to language: the declarative/procedural model. *Cognition* 92, 231–270. doi: 10.1016/j.cognition.2003.10.008
- Vallet, G. T. (2015). Embodied cognition of aging. *Front. Psychol.* 6:463. doi: 10.3389/fpsyg.2015.00463
- van Heuven, W. J. B., and Dijkstra, T. (2010). Language comprehension in the bilingual brain: fMRI and ERP support for psycholinguistic models. *Brain Res. Rev.* 64, 104–122. doi: 10.1016/j.brainresrev.2010.03.002
- Vukovic, N., Feurra, M., Shpektor, A., Myachykov, A., and Shtyrov, Y. (2017). Primary motor cortex functionally contributes to language comprehension: an online rTMS study. *Neuropsychologia* 96, 222–229. doi: 10.1016/j.neuropsychologia.2017.01.025
- Vukovic, N., and Shtyrov, Y. (2014). Cortical motor systems are involved in second-language comprehension: evidence from rapid mu-rhythm desynchronization. *Neuroimage* 102(Part 2), 695–703. doi: 10.1016/j.neuroimage.2014.08.039
- Vukovic, N., and Williams, J. N. (2014). Automatic perceptual simulation of first language meanings during second language sentence processing in bilinguals. *Acta Psychol.* 145(Suppl. C), 98–103. doi: 10.1016/j.actpsy.2013.11.002
- Warrington, E. K., and McCarthy, R. A. (1987). Categories of knowledge: further fractionations and an attempted integration. *Brain* 110, 1273–1296. doi: 10.1093/brain/110.5.1273
- Wartenburger, I., Heekeren, H. R., Abutalebi, J., Cappa, S. F., Villringer, A., and Perani, D. (2003). Early setting of grammatical processing in the bilingual brain. *Neuron* 37, 159–170. doi: 10.1016/S0896-6273(02)01150-9
- Wellsby, M., and Pexman, P. M. (2014). Developing embodied cognition: insights from children's concepts and language processing. *Front. Psychol.* 5:506. doi: 10.3389/fpsyg.2014.00506
- Willems, R. M., Labruna, L., D'Esposito, M., Ivry, R., and Casasanto, D. (2011). A functional role for the motor system in language understanding: evidence from theta-burst transcranial magnetic stimulation. *Psychol. Sci.* 22, 849–854. doi: 10.1177/0956797611412387
- Willems, R. M., Toni, I., Hagoort, P., and Casasanto, D. (2009). Neural dissociations between action verb understanding and motor imagery. *J. Cogn. Neurosci.* 22, 2387–2400. doi: 10.1162/jocn.2009.21386
- Xue, J., Marmolejo-Ramos, F., and Pei, X. (2015). The linguistic context effects on the processing of body-object interaction words: an ERP study on second language learners. *Brain Res.* 1613(Suppl. C), 37–48. doi: 10.1016/j.brainres.2015.03.050
- Zanini, S., Tavano, A., and Fabbro, F. (2010). Spontaneous language production in bilingual Parkinson's disease patients: evidence of greater phonological, morphological and syntactic impairments in native language. *Brain Lang.* 113, 84–89. doi: 10.1016/j.bandl.2010.01.005
- Zanini, S., Tavano, A., Vorano, L., Schiavo, F., Gigli, G. L., Aglioti, S. M., et al. (2004). Greater syntactic impairments in native language in bilingual Parkinsonian patients. *J. Neurol. Neurosurg. Psychiatry* 75, 1678–1681. doi: 10.1136/jnnp.2003.018507
- Zawiszewski, A., Gutiérrez, E., Fernández, B., and Laka, I. (2011). Language distance and non-native syntactic processing: evidence from event-related potentials*. *Biling. Lang. Cogn.* 14, 400–411. doi: 10.1017/S1366728910000350
- Zhang, S., and Thompson, N. (2004). DIALANG: a diagnostic language assessment system (review). *Can. Mod. Lang. Rev.* 61, 290–293. doi: 10.1353/cml.2005.0011
- Zimmer, H. T. (2001). "Why do actions speak louder than words?: action memory as a variant of encoding manipulations or the result of a specific memory system," in *Memory Far Action: A Distinct Form of Episodic Memory?*, eds H. D. Zimmer and R. L. Cohen (Oxford: Oxford University Press).
- Zwaan, R. A., Taylor, L. J., and de Boer, M. (2010). Motor resonance as a function of narrative time: further tests of the linguistic focus hypothesis. *Brain Lang.* 112, 143–149. doi: 10.1016/j.bandl.2008.11.004

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Monaco, Jost, Gygax and Annoni. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Female Speakers Benefit More Than Male Speakers From Prosodic Charisma Training—A Before-After Analysis of 12-Weeks and 4-h Courses

Oliver Niebuhr^{1*}, Silke Tegtmeier² and Tim Schweisfurth²

¹ Centre for Industrial Electronics, Mads Clausen Institute, University of Southern Denmark, Sønderborg, Denmark,

² Department of Technology Entrepreneurship and Innovation, Mads Clausen Institute, University of Southern Denmark, Sønderborg, Denmark

OPEN ACCESS

Edited by:

Yury Y. Shtyrov,
Aarhus University, Denmark

Reviewed by:

Silke Paulmann,
University of Essex, United Kingdom
Francesca D'Errico,
Roma Tre University, Italy

*Correspondence:

Oliver Niebuhr
olni@sdu.dk

Specialty section:

This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Communication

Received: 27 September 2018

Accepted: 13 March 2019

Published: 03 April 2019

Citation:

Niebuhr O, Tegtmeier S and
Schweisfurth T (2019) Female
Speakers Benefit More Than Male
Speakers From Prosodic Charisma
Training—A Before-After Analysis of
12-Weeks and 4-h Courses.
Front. Commun. 4:12.
doi: 10.3389/fcomm.2019.00012

Perceived charisma is an important success factor in professional life. However, women are worse than men in conveying physical charisma signals while at the same time having to perform better than men in order to be perceived equally charismatic. Speech prosody probably contains the most influential charisma signals. We have developed a system called “Pascal” that analyzes and assesses on objective acoustic grounds how well-speakers employ their prosodic charisma parameters. Pascal is used for charismatic-speech training in 12-weeks and 4-h courses on entrepreneurship and leadership. Comparing the prosodic-charisma scores for a total of 72 participants at the beginning and end of these two course types showed that female speakers start with significantly lower prosodic-charisma scores than male speakers. However, at the end of the 4-h course, female speakers can catch up with their male counterparts in terms of prosodic charisma. At the end of the 12-weeks courses, male speakers keep their lead, but female speakers are able to significantly reduce the prosodic charisma gap to male speakers. Since leadership and entrepreneurship are still male-dominated domains, our results can be seen as an encouragement for women to attend prosodic charisma training. Furthermore, these courses require a gender-specific design as we found men to improve mainly in F0 parameters and women in duration and phonation parameters.

Keywords: prosody, charisma, entrepreneurship, rhetoric, sex differences, phonetics, speaker training, leadership

INTRODUCTION

Charismatic speakers convey emotionally “contagious” (Fox Cabane, 2012, p. 145) verbal and non-verbal signals that make others invest their thoughts, actions, time or money into them (Antonakis et al., 2016). There seems to be a tangible cognitive reason for these charisma effects: Perceived charisma can inhibit areas of the brain that are associated with cognitive control of behavior and abstract reasoning (Schjødt et al., 2010). Acoustic cues alone predict perceived speaker charisma with an accuracy of 66–75% (Chen et al., 2014; Park et al., 2014), with prosodic features making the largest contribution to this prediction accuracy (cf. also Gregory and Gallagher, 2002). It was findings like these that recently boosted the interest of phoneticians in charismatic speech.

Phonetic studies have shown that acoustic-prosodic parameters are positively correlated with perceived speaker charisma (Strangert and Gustafson, 2008; Rosenberg and Hirschberg, 2009; Signorello et al., 2012; D'Errico et al., 2013; Niebuhr et al., 2016, 2018). For example, higher F0 levels, larger F0 ranges, and faster speaking rates make speakers sound more charismatic. The same also applies to the intensity level and standard deviation as well as to the frequency of emphatically stressed words. Disfluency count (incl. silent pauses), formant levels (F1–F3), prosodic-phrase durations, and spectral-slope voice-quality measures are negatively correlated with perceived speaker charisma.

Women sound less charismatic than men (Brands et al., 2015; Jokisch et al., 2018). This is true even if all other factors besides gender (incl. all prosodic parameters) are equal (Novák-Tót et al., 2017; Niebuhr et al., 2018). This is rarely the case in everyday life because men's speaking skills are promoted and valued more and perhaps even judged less critically by the society than those of women (Baxter, 1999; Sellnow and Treinen, 2004; Cameron, 2006).

Charisma also plays a big role in entrepreneurship; both with respect to the success of young companies in being innovative (Todorovic and Schlosser, 2007) and the success of individuals in founding a sustainable start-up business through legitimizing and fund-raising activities (Clark, 2008; Davis et al., 2017). The relevance of charisma in entrepreneurship and the finding that women sound less charismatic than men together resonate with what is known as the “gender gap” in entrepreneurship (Markussen and Røed, 2017). Besides the fact that many investors think of entrepreneurship as a male domain (Marlow, 2014) and, therefore, “screen out” women's business ideas while men's ideas are “screened in” (Kanze et al., 2018), women's oral presentations in front of investors sound less persuasive and are less likely to be funded and supported than those of men (Brooks et al., 2014).

A male advantage also exists in politics, although perhaps less strongly so than in entrepreneurship (Bystorm et al., 2001). Additionally, entrepreneurship represents a pillar of economic growth in today's innovation-driven economies (Audretsch et al., 2006), with female entrepreneurs making a disproportionate contribution to this growth (Gutierrez, 2017). Therefore, the question arises as to how this unfavorable gender gap in entrepreneurship can be closed. One answer is: with an effective way of turning female entrepreneurs into more charismatic speakers. The growing phonetic understanding of perceived speaker charisma has enabled researchers to develop computer-based systems for a precise parametric assessment and objective training of charismatic speech. Some systems are multi-modal, like Cicero (Batrinsa et al., 2013) and MACH (Hoque et al., 2013). Our system focuses on the acoustic key parameters of prosody. It is called “Pascal”: Prosodic Analysis of Speaker Charisma—Assessment and Learning.

While users' learning success has already been documented for Cicero and MACH (Batrinsa et al., 2013; Hoque et al., 2013), it still needs to be checked for Pascal. Moreover, neither system has been analyzed so far as to whether men or women benefit differently from using it. There is growing evidence that women are more sensitive to emotional and interactional

prosodic elements than men, and that they also use these elements to a greater extent in speech production (Daly and Warren, 2001; Haan, 2002; Lausen and Schacht, 2018). So, if Pascal is able to shift a speaker's prosodic parameters in a more charismatic direction, it is possible that women benefit more from Pascal training than men, meaning that Pascal training would be a suitable means to reduce or close the “gender gap” in entrepreneurship. Specifically, the following questions are addressed: (1) Do women have a lower baseline (untrained) prosodic charisma level than men when they perform an entrepreneurial task like giving a short investor-oriented presentation of a business idea? (2) Is prosodic training able to shift a speaker's speech parameters in a more charismatic direction? (3) Do men and women benefit equally from such training or is there a gender specificity? (4) How are (2) and (3) affected by training time, i.e., in the comparison of a short crash course and a long intensive course?

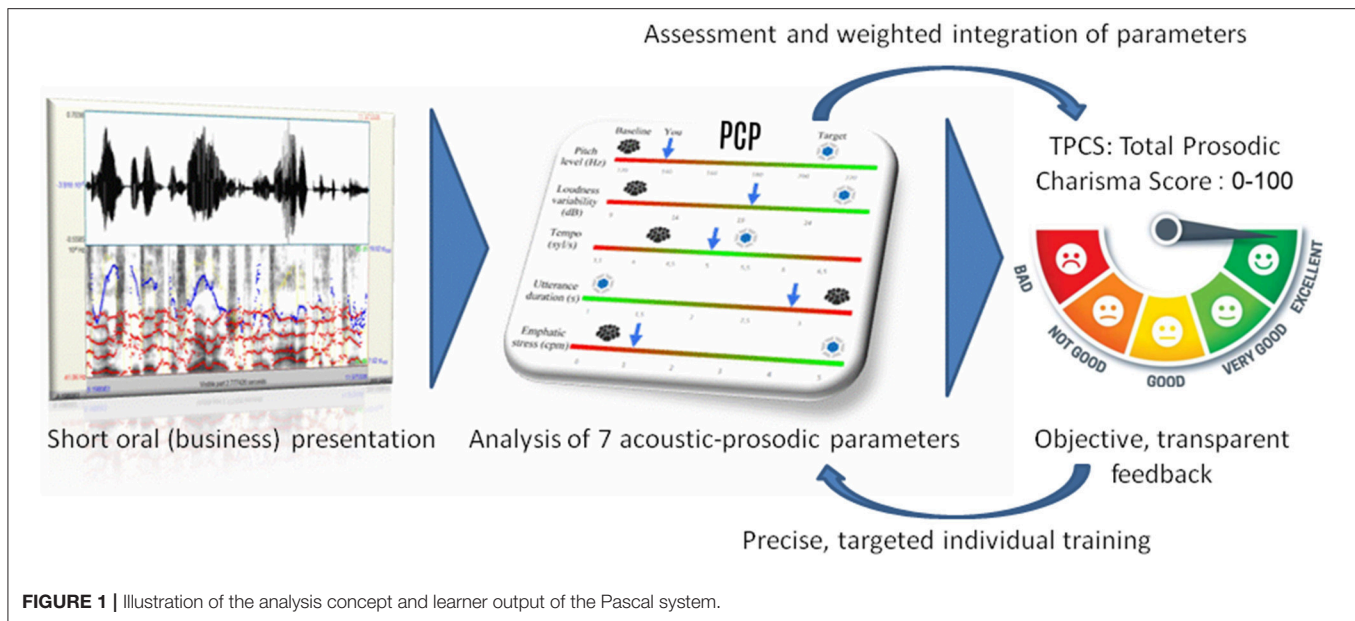
METHODS

Pascal and TPCS

Pascal is the patent-pending result of years of experimental-phonetic research in speech production and perception. It is based on the correlations listed in the Introduction between prosodic parameters and perceived speaker charisma, with one crucial innovation: the notion of an “overdose” (cf. also Rosenberg and Hirschberg, 2009). Perceived speaker charisma cannot be infinitely increased by a prosodic parameter shift in a given direction. Above a certain “overdose” threshold, the effect is reversed, making a further parameter shift increasingly detrimental for perceived speaker charisma. For example, a higher F0 level and a faster speaking rate make a speaker sound more charismatic, but when speakers get too high-pitched or too fast, the charisma level drops drastically. These “overdose” thresholds have been determined for each parameter in a large-scale series of perception experiments, also taking into account confounding variables like speaker gender. Furthermore, by playing off each parameter against the others in these perception experiments, we determined perceptual weights for the individual prosodic parameters (Berger et al., 2017). For example, after having found that F0 range is more relevant for perceived speaker charisma than F0 level, we developed multipliers for both F0 parameters that appropriately represent their relevance difference in Pascal's user feedback. Further details of Pascal are outlined in Niebuhr et al. (2017).

On this basis, any recorded speech sample can be uploaded to Pascal. The system then breaks down the sample into its relevant prosodic parameters, determines the mean parameter levels, and returns a twofold output: a Total Prosodic Charisma Score (TPCS), and a user-friendly Prosodic Charisma Profile (PCP), showing how the speaker performs on each prosodic parameter in relation to the overdose thresholds (red sections on the PCP scales), see **Figure 1**.

The TPCS is the dependent variable of the present paper. Using the TPCS as dependent variable does not mean that subjective charisma performances are compared. The TPCS is primarily rooted in acoustics. It translates acoustic-prosodic



parameter values into a psychoacoustic measure that is calibrated through listener ratings. In this respect, TPCS is similar to the translation of F0 (Hz) into perceived pitch along the Mel scale or the translation of acoustic energy (dB) into perceived loudness along the Sone scale (Fastl and Zwicker, 2006). However, Mel and Sone are both scalar psychoacoustic translations of single acoustic parameters. Other parameters have to be constant when measuring Mel and Sone (e.g., 1 kHz, 1 s, pure tone etc.). In contrast, the TPCS integrates multiple acoustic parameters and offsets them against each other. In addition, the TPCS is, through its listener ratings, already normalized for the gender bias in perceived speaker charisma. Thus, a man and a woman both with a TPCS of, for example, “35” can be assumed to convey equally strong prosodic charisma signals. Of course, other speaker factors like foreign accent, body language, verbal rhetoric, physical attractiveness etc. can still make these (and any other) two speakers differently charismatic overall for listeners (e.g., Antonakis et al., 2011; Fox Cabane, 2012; Scherer et al., 2012). However, the present study is exclusively concerned with prosody and gender-specific differences in TPCS levels and improvements. These differences and improvements matter irrespective of all other possible sources of charisma that are not taken into account and controlled here. Moreover, these other factors are irrelevant here insofar as they have no influence on the acoustic TPCS measurements.

Charisma Training Courses

In entrepreneurship education, Pascal has been used regularly since 2017 in two different types of courses whose participants learn how to give successful business presentations. One course is a long intensive course that consists of 12 lectures of 90 min each over a whole semester. The other course is a short crash course of 4-h on a single day. Both courses are for entrepreneurs with an academic background in business engineering who plan to found a new start-up company or lead an innovation

department within an existing company. The 12-weeks and 4-h courses obviously differ in the amount and detail of multi-modal rhetorical information provided to participants. However, at the heart of both courses is the successive improvement of prosodic charisma parameters through the reiterative use of Pascal. Participants upload their speech samples to the system, receive automatic PCP and TPCS feedback, interpret this feedback, and then try to produce a more charismatic speech. The type of speech was in all cases a short oral business presentation of 3–5 min, given in L2 English. The presentation was uploaded to and analyzed by Pascal as a whole.

Crucially, there is no systematically different treatment of men and women in the courses. Firstly, the primary feedback comes from the unbiased Pascal system itself. The human course instructor (first author) only adds explanations, clarifications, and guidance to the machine feedback. Secondly, the present paper is a *post-hoc* analysis of existing data. At the time of the courses, the course instructor did not yet know the gender-specific questions of the paper and, thus, could not influence participants consciously or subconsciously.

The 12-weeks and 4-h courses both start with the participants holding their business-idea presentation for the first time; and the two courses both end with a final matured presentation (of the same business idea) in which each participant can showcase what s/he has learned. Both presentations are fed into Pascal. The two resulting PCPs represent the baseline profile (BP) and the trained profile (TP). Their corresponding TPCSs are compared here. The 12-weeks and 4-h courses are entirely given in English, which includes the business-idea presentations as well as the expert supervision and Pascal’s PCP and TPCS feedback interface.

Course Participants

The 12-weeks intensive course included 35 participants, 20 males, and 15 females. The 4-h crash course was carried out with 37 participants, 21 males and 16 females. All participants

held, after about 15 years of education, an academic degree in business engineering or management. Moreover, all participants were post-graduate university students with part-time company employment and at least 1 year of working experience. The proficiency level of L2 English was at least B2, according to university-internal aptitude tests. There were no English native speakers in the courses.

Native-language background differed between the course participants. In the 12-weeks course, the majority of the 20 male speakers had German as their native language (35%), followed by Danish (25%), Slavic languages like Russian or Czech (20%), Arabic (15%), and Mandarin Chinese (5%). The percentages were similar for the 15 female speakers in that course (German 40%, Danish 20%, Slavic languages 20%, Arabic 6.7%, Mandarin Chinese 13.3%). In the 4-h course, the majority of speakers were again German (52.4% in the male and 56.3% in the female speaker sample). The percentages of Scandinavian speakers, Slavic speakers, Arabic speakers, and Mandarin Chinese speakers were all similarly low (between 1.9 and 6.3%), but in all cases larger than 0%. Chi-squared tests were carried out and showed by lack of significant differences that the native-language backgrounds were similarly distributed between the male and female speaker samples of each course. The same applied to speaker age. It ranged from 21 to 31 years in the 12-weeks course ($\bar{\mu} = 23.6$ years, $sd = 2.6$ years) and from 22 to 45 years in the 4-h course ($\bar{\mu} = 28.8$ years, $sd = 3.2$ years).

An ethics approval was not required for this empirical but non-experimental research as per institutional guidelines and national regulations. We adhered to the Danish Code of Conduct for Research Integrity and current data protection rules (GDPR). All our course participants gave informed written consent (cf. Declaration of Helsinki) that their data is recorded and can be used for scientific analysis in an entirely anonymous fashion.

RESULTS

A three-way General Linear Mixed Model (GLMM) analysis was conducted for the TPCS data (dependent variable). Course Type (12 weeks vs. 4-h) and Gender (male vs. female) were between-subject factors, and Training (BP vs. TP) was a within-subject factor. Individual speaker was included as a covariate. A descriptive results summary of BP and TP values across the three factors is shown in **Figure 2**.

The three-way GLMM yielded a significant main effect of Gender [$F_{(1,68)} = 26.19$, $p < 0.001$] as the acoustic-prosodic charisma scores (TP and BP) were on average significantly lower for female than for male speakers in our courses [$\bar{\mu}$ males = 59.9 vs. $\bar{\mu}$ females = 49.8; $t_{(81,61)} = 2.84$, $p = 0.002$]. There is also a significant main effect of Training [$F_{(1,68)} = 268.62$, $p < 0.001$] as TP scores were, across all speakers and both courses, higher than BP scores, thus indicating the participants' significant TPCS improvement from the beginning to the end of a course [$\bar{\mu}$ BP_{m&f} = 40.0 vs. $\bar{\mu}$ TP_{m&f} = 71.2; $t_{(71)} = -12.31$, $p < 0.001$]. A significant interaction between Training and Gender [$F_{(1,68)} = 9.92$, $p = 0.002$] reflects that women's BP scores were a lot lower than men's BP scores ($\bar{\mu}$ BP_m = 46.9 vs. $\bar{\mu}$ BP_f = 30.7), but that this

score difference became smaller (but was overall still significant) in the TP recordings ($\bar{\mu}$ TP_m = 72.8 vs. $\bar{\mu}$ TP_f = 69.1; $t_{(40,30)} = 2.03$, $p = 0.04$).

Course Type had no separate significant main effect. However, we found significant interactions between Course Type and Training [$F_{(1,68)} = 36.61$, $p < 0.001$] and Course Type and Gender [$F_{(1,68)} = 7.55$, $p = 0.03$]. The three-way interaction was not significant.

In order to examine the effect of Course Type in more detail, we split up the three-way GLMM and ran two additional separate two-way GLMMs, one on the 12-weeks intensive course and one on the 4-h crash course. The additional GLMMs replicated, separately for both courses, the beneficial effect of Training on TP scores [12-weeks: $F_{(1,66)} = 189.53$, $p < 0.001$; 4-h: $F_{(1,64)} = 148.95$, $p < 0.001$]. However, the size of this Training effect, i.e., the learning success in terms of the speakers' acoustic-prosodic charisma improvement, differed depending on Course Type. The improvement was larger in the 12-weeks intensive course than in the 4-h crash course, which caused the significant interaction Course Type*Training in the three-way GLMM.

The two additional GLMMs also replicated the significant interaction of Training and Gender for both courses [12-weeks: $F_{(1,66)} = 16.04$, $p < 0.001$; 4-h: $F_{(1,64)} = 14.96$, $p < 0.001$]. Again, these interactions differed depending on Course Type. In the 12-weeks course, the interaction reflects that women managed to significantly reduce men's TPCS lead from BP to TP recordings [$\bar{\mu}$ BP_m = 41.1 vs. $\bar{\mu}$ BP_f = 24.5; $t_{(19,14)} = 3.67$, $p < 0.001$; $\bar{\mu}$ TP_m = 81.7 vs. $\bar{\mu}$ TP_f = 72.5; $t_{(19,14)} = 2.00$, $p = 0.027$]. In the 4-h course, women were even able to turn their lower initial BP performance [$\bar{\mu}$ BP_m = 52.6 vs. $\bar{\mu}$ BP_f = 36.9; $t_{(20,15)} = 4.42$, $p < 0.001$] into a TP performance at eye level with men ($\bar{\mu}$ TP_m = 64.3 vs. $\bar{\mu}$ TP_f = 66.0; n.s.). It was for this reason that the interaction Course Type*Gender became significant and that the women's overall disadvantage over men in terms of TPCSs only showed up as a significant main effect in the GLMM on the 12-weeks course [$F_{(1,64)} = 7.65$, $p = 0.008$].

DISCUSSION

Our results suggest positive answers to all questions raised in the Introduction. (1) Women have a lower baseline prosodic charisma level (BP) than men when they perform an entrepreneurial task like giving a short investor-oriented presentation of a business idea. This is true in the same way for both independent speaker samples of the 4-h and 12-weeks courses. Our supporting evidence on question (1) matches with the well-known "gender gap" in entrepreneurship, particularly since the speaking task was an entrepreneurial one and all analyzed speakers are in some way involved in entrepreneurial activities. However, the "gender gap" in entrepreneurship underlies a complex explanation (Markussen and Røed, 2017), and the prosodic charisma gap between prosodically untrained men and women can only be one component of its origin.

Furthermore, we found clear empirical evidence that (2) acoustically based prosody training is able to shift a speaker's prosodic parameters in a more charismatic direction and that,

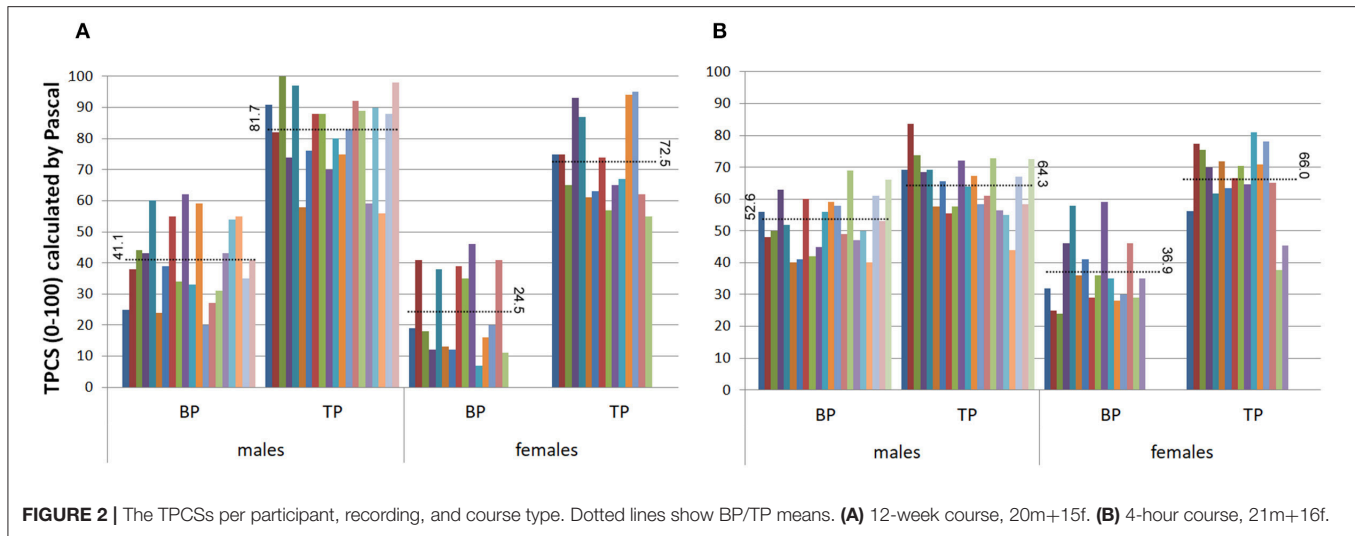


FIGURE 2 | The TPCSs per participant, recording, and course type. Dotted lines show BP/TP means. **(A)** 12-week course, 20m+15f. **(B)** 4-hour course, 21m+16f.

most importantly, (3) there is a gender specificity in the effect of (2). Women do benefit more from this prosody training than men. Compared to men, women's prosodic charisma improved *faster* so that they caught up with male-speaker performances after only 4-h of training. In combination with the finding that women started at lower BP-TPCS levels, this also means that women learn *more* during acoustically based prosody training. Men could only maintain a small TPCS lead after a longer period of training, as in the 12-weeks intensive course. Thus, (4) training time has an effect.

Note that decomposing the TPCS into its individual parameters revealed that the improvements of men and women rely partly on different parameters. Men improved mainly through higher F0 levels ranges and more emphatically-stressed words. Women's improvement was primarily based on shorter prosodic-phrase durations and better intensity and voice-quality measurements. This difference applies to the speaker samples of both courses. In combination with the lack of significant demographic differences between the male and female speakers within both courses, the findings suggest that prosodic charisma training needs to be gender-tailored in order to be effective

across the full range of acoustic-prosodic parameters. However, our speaker sample is relatively small, and the sample sizes are not balanced across gender groups. Therefore, it is not clear how far these findings can be generalized, particularly beyond the analyzed speakers' age group and educational and socio-economic backgrounds as well as beyond other settings than short business presentations held in non-native English. These aspects of generalization are obvious starting points for follow-up studies. We are currently focusing on the aspect of native-language background.

Nevertheless, taken together, the present findings should be seen as a strong encouragement for female entrepreneurs to take part in a prosodic charisma training. It can make a significant contribution to counterbalance the stronger financial and technical support of male business ideas by male investors or decision makers.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

REFERENCES

- Antonakis, J., Bastardo, N., and Jacquart, P. (2016). Charisma: an ill-defined and ill-measured gift. *Ann. Rev. Organ. Psychol. Organ. Behav.* 3, 293–319. doi: 10.1146/annurev-orgpsych-041015-062305
- Antonakis, J., Fenley, M., and Liechti, S. (2011). Can charisma be taught? Tests of two interventions. *Acad. Manag. Learn. Educ.* 10, 374–396. doi: 10.5465/amle.2010.0012
- Audretsch, D. B., Keilbach, M. C., and Lehmann, E. (2006). *Entrepreneurship and Economic Growth*. Oxford, NY: Oxford University Press. doi: 10.1093/acprof:oso/9780195183511.001.0001
- Batrinca, L., Stratou, G., Shapiro, A., Morency, L.-P., and Scherer, S. (2013). "Cicero - towards a multimodal virtual audience platform for public speaking training," in *Proceedings Intelligent Virtual Agents 2013* (Edinburgh).
- Baxter, J. A. (1999). Teaching girls to speak out: the female voice in public contexts. *Lang. Educ.* 13, 81–98.
- Berger, S., Niebuhr, O., and Peters, B. (2017). "Winning over an audience - a perception-based analysis of prosodic features of charismatic speech," in *Proceedings 43rd Annual Meeting of the German Acoustical Society* (Kiel), 793–796.
- Brands, R. A., Menges, J. I., and Kilduff, M. (2015). The leader-in-social-network schema: perceptions of network structure affect gendered attributions of charisma. *Org. Sci.* 26, 1210–1225. doi: 10.1287/orsc.2015.0965
- Brooks, A. W., Huang, L., Kearney, S. W., and Murray, F. E. (2014). Investors prefer entrepreneurial ventures pitched by attractive men. *Proc. Natl. Acad. Sci. U.S.A.* 111, 4427–4431. doi: 10.1073/pnas.1321202111
- Bystrom, D. G., Robertson, T. A., and Banwart, M. C. (2001). Framing the fight - An analysis of media coverage of male and female candidates in primary races for Governor and U.S. Senate in 2000. *Am. Behav. Sci.* 44, 1999–2013. doi: 10.1177/00027640121958456
- Cameron, D. (2006). "Theorising the female voice in public context," in *Speaking Out*, ed J. Baxter (New York, NY: Palgrave), 3–20.

- Chen, L., Feng, G., Joe, J., Leong, C. W., Kitchen, C., and Lee, C. M. (2014). "Towards automated assessment of public speaking skills using multimodal cues," in *Proceedings 16th International Conference on Multimodal Interaction* (Istanbul).
- Clark, C. (2008). The impact of entrepreneurs' oral 'pitch' presentation skills on business angels' initial screening investment decisions. *Vent. Cap.* 10:3, 257–279. doi: 10.1080/13691060802151945
- Daly, N., and Warren, P. (2001). Pitching it differently in New Zealand english: speaker sex and intonation patterns. *J. Sociol.* 5, 85–96. doi: 10.1111/1467-9481.00139
- Davis, B. C., Hmieleski, K. M., Webb, J. W., and Coombs, J. E. (2017). Funders' positive affective reactions to entrepreneurs' crowd-funding pitches: the influence of perceived product creativity and entrepreneurial passion. *J. Bus Vent.* 32, 90–106. doi: 10.1016/j.jbusvent.2016.10.006
- D'Errico, F., Signorello, R., Demolin, D., and Poggi, I. (2013). "The perception of charisma from voice. A crosscultural study," in *Proceedings Humaine Association Conference on Affective Computing and Intelligent Interaction* (Geneva).
- Fastl, H., and Zwicker, E. (2006). *Psychoacoustics - Facts and Models*. Berlin; Heidelberg: Springer.
- Fox Cabane, O. (2012). *The Charisma Myth: How Anyone Can Master the Art and Science of Personal Magnetism*. New York, NY: Penguin.
- Gregory, S. W. Jr., and Gallagher, T. J. (2002). Spectral analysis of candidates' nonverbal vocal communication: predicting U.S. presidential election outcomes. *Soc. Psychol. Q.* 65, 298–308. doi: 10.2307/3090125
- Gutierrez, C. M. (2017). *Women Entrepreneurs Are Driving Economic Growth*. Available online at: https://www.huffingtonpost.com/entry/women-entrepreneurs-are-driving-economic-growth_us_59f7c3dce4b04494283378f3
- Haan, J. (2002). *Speaking of Questions*. Utrecht: LOT.
- Hoque, M. E., Courgeon, M., Martin, J.-C., Mutlu, B., and Picard, R. W. (2013). "MACH: my automated conversation coach," in *Proceedings ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Zurich).
- Jokisch, O., Iaroshenko, V., Maruschke, M., and Ding, H. (2018). "Influence of Age, gender and sample duration on the charisma assessment of German speakers," in *Proceedings 29th Conference on Electronic Speech Signal Process* (Ulm).
- Kanze, D., Huang, L., Conley, M. A., and Higgins, E. T. (2018). We ask men to win & women not to lose: closing the gender gap in startup funding. *Acad. Manag. J.* 61, 1–29. doi: 10.5465/amj.2016.1215
- Lausen, A., and Schacht, A. (2018). Gender differences in the recognition of vocal emotions. *Front. Psychol.* 9:882. doi: 10.3389/fpsyg.2018.00882
- Markussen, S., and Røed, K. (2017). The gender gap in entrepreneurship – The role of peer effects. *J. Econom. Behav. Organ.* 134, 356–373. doi: 10.1016/j.jebo.2016.12.013
- Marlow, S. (2014). Exploring future research agendas in the field of gender and entrepreneurship. *Int. J. Gender Entrepr.* 6, 102–120. doi: 10.1108/IJGE-01-2013-0003
- Niebuhr, O., Skarnitzl, R., and Tylečková, L. (2018). "The acoustic fingerprint of a charismatic voice - Initial evidence from correlations between long-term spectral features and listener ratings," in *Proceedings 18th International Conference of Speech Prosody* (Poznań). doi: 10.21437/SpeechProsody.2018-73
- Niebuhr, O., Tegtmeier, S., and Brem, A. (2017). Advancing research and practice in entrepreneurship through speech analysis – from descriptive rhetorical terms to phonetically informed acoustic charisma metrics. *J. Speech Sci.* 6, 3–26.
- Niebuhr, O., Voße, J., and Brem, A. (2016). What makes a charismatic speaker? A computer-based acoustic prosodic analysis of Steve Jobs tone of voice. *Comput. Hum. Behav.* 64, 366–382. doi: 10.1016/j.chb.2016.06.059
- Novák-Tót, E., Niebuhr, O., and Chen, A. (2017). "A gender bias in the acoustic-melodic features of charismatic speech?" in *Proceedings Annual Conference of the International Speech Communication Association* (Stockholm), 2248–2252. doi: 10.21437/Interspeech.2017-1349
- Park, S., Shoemark, P., and Morency, L.-P. (2014). "Toward crowd-sourcing micro-level behavior annotations: the challenges of interface, training, and generalization," in *Proceedings of the 18th International Conference on Intelligent User Interfaces* (Santa Monica, CA).
- Rosenberg, A., and Hirschberg, J. (2009). Charisma perception from text and speech. *Speech Commun.* 51, 640–655. doi: 10.1016/j.specom.2008.11.001
- Scherer, S., Layher, G., Kane, J., Neumann, H., and Campbell, N. (2012). "An audiovisual political speech analysis incorporating eye-tracking and perception data," in *Proceedings 8th International Conference on Language Resources and Evaluation* (LREC'12) (Istanbul).
- Schjød, H., Stodkilde-Jørgensen, A. W., Geertz, T. E., and Lund, A. (2010). roepstorffthe power of charisma-perceived charisma inhibits the frontal executive network of believers in intercessory prayer social cognitive and affective. *Neuroscience.* 6, 119–127. doi: 10.1093/scan/nsq023
- Sellnow, D. D., and Treinen, K. P. (2004). The role of gender in perceived speaker competence: an analysis of student peer critiques. *Commun. Educ.* 53, 286–296. doi: 10.1080/0363452042000265215
- Signorello, R., D'Errico, F., Poggi, I., and Demolin, D. (2012). "How charisma is perceived from speech: a multidimensional approach. privacy, security, risk and trust (PASSAT)," in *International Conference on Social Computing (SocialCom)* (Amsterdam).
- Strangert, E., and Gustafson, J. (2008). "What makes a good speaker? Subject ratings, acoustic measurements and perceptual evaluations," in *Proceedings of the 9th International Interspeech Conference* (Brisbane), 1688–1691.
- Todorovic, W., and Schlosser, F. K. (2007). An entrepreneur and a leader! a framework conceptualizing the influence of leadership style on a firm's entrepreneurial orientation-performance relationship. *J. Small Business Entrepr.* 20, 289–307. doi: 10.1080/08276331.2007.10593401

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Niebuhr, Tegtmeier and Schweisfurth. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Explicitly Slow, Implicitly Fast, or the Other Way Around? Brain Mechanisms for Word Acquisition

Yury Shtyrov^{1,2*}, Alexander Kirsanov^{2,3} and Olga Shcherbakova^{2,3}

¹ Department of Clinical Medicine, Center of Functionally Integrative Neuroscience, Aarhus University, Aarhus, Denmark, ² Laboratory of Behavioural Neurodynamics, Saint Petersburg State University, Saint Petersburg, Russia, ³ Department of General Psychology, Faculty of Psychology, Saint Petersburg State University, Saint Petersburg, Russia

Keywords: learning, memory, fast mapping, word, brain, neocortex, neuroimaging, explicit encoding

Our ability to communicate using language is a specific cognitive faculty that makes humans stand apart from all other animal species on the planet. Despite the crucial role that language plays in our individual and social well-being, the origins of language are still poorly understood from both evolutionary and ontogenetic perspectives. One of the key gaps in the knowledge lies in the understanding of specific cognitive and neural bases of language acquisition that underpin our successful and efficient ability to learn a large number of new words, both as children at all stages of development and as adults when learning a new language or novel professional lexicon. This opinion paper briefly overviews the main systems involved in word acquisition, identifies gaps in the existing evidence and suggests possible ways to close them.

The behavioral and neural mechanisms of word acquisition remain a debated topic (for reviews, see e.g., Dollaghan, 1985; Davis and Gaskell, 2009). On the systems level, learning processes are most commonly separated into initial encoding and later consolidation. The stage of encoding is believed to occur rapidly and to involve multiple brain areas, most crucially medial temporal lobe (MTL) including hippocampus and parahippocampal cortices (McClelland et al., 1995; Suzuki, 2006); the consolidation, in turn, is a more gradual process leading to the formation of long-term memory traces in the neocortex (Walker and Stickgold, 2006; Battaglia et al., 2011). Such a two-stage or “complementary learning systems” approach resonates through different levels of investigations, including animal studies with hippocampal and cortical lesions trying to disentangle the two stages (Talpos et al., 2008), cognitive science models using computational neural networks to simulate neural memory build-up processes (O’Reilly and McClelland, 1994), as well as patient studies using hippocampally-damaged amnesiacs that demonstrate specific patterns of retrograde memory loss (Scoville and Milner, 1957; Sharon et al., 2011). A range of experiments extended this approach to account for the brain’s word learning mechanisms, with their results indicating that newly-learned word-forms fully enter the lexicon only after an overnight consolidation period, which is accompanied by changes in neocortical and MTL activity (Gaskell and Dumay, 2003; Davis and Gaskell, 2009). While this framework can successfully explain a range of phenomena in the fields of memory, learning, and language, another body of observations suggests the existence of a hippocampally independent route for direct acquisition of new word forms by the neocortex, at least under certain conditions (Shtyrov, 2012), as we will discuss below.

Whereas different processes (e.g., imitation, repetition, association, or generalization) may be involved in learning, the initial acquisition of new words in real-life situations can arguably be achieved through two main learning strategies: a direct explicit instruction (e.g., “This is a glorp, please remember it”) or a contextually-driven implicit inference/deduction (“There is a toy car, a book and a glorp on the table. Which color is the glorp?”). Although not mutually exclusive, these two are characterized by dissociable (to a degree at least) properties. Explicit learning, often dubbed explicit encoding (EE), is usually associated with repetitive presentation occurring over extended (or even multiple) practice sessions, such as classroom instruction or rehearsal.

OPEN ACCESS

Edited by:

S. H. Annabel Chen,
Nanyang Technological University,
Singapore

Reviewed by:

Chiao-Yi Wu,
Nanyang Technological University,
Singapore

Adolfo M. Garcia,
Laboratory of Experimental
Psychology and Neuroscience,
Argentina

*Correspondence:

Yury Shtyrov
yury@cfin.au.dk

Received: 06 November 2018

Accepted: 19 March 2019

Published: 26 April 2019

Citation:

Shtyrov Y, Kirsanov A and
Shcherbakova O (2019) Explicitly
Slow, Implicitly Fast, or the Other Way
Around? Brain Mechanisms for Word
Acquisition.

Front. Hum. Neurosci. 13:116.
doi: 10.3389/fnhum.2019.00116

In contrast, contextually-driven deduction normally takes place in routine daily interactions between individuals and appears to have a near-immediate effect, evident before long-term memory consolidation processes set in. For building up new semantic representations, it requires just a few expositions (with claims of even single-shot learning) in a context that facilitates inference through exclusion or deduction (Bloom and Markson, 1998; Halberda, 2006; Horst and Samuelson, 2008). This rapid implicit acquisition is often termed fast mapping (FM) and is considered to be a general learning mechanism that plays a key role in acquiring new words and their semantics in the process of natural language learning (Carey and Bartlett, 1978; Kaminski et al., 2004).

Even though, as discussed below, teasing the two mechanisms apart is not straightforward, it is this latter strategy, FM, which has been argued to predominantly depend on the neocortex and be largely independent from the MTL and hippocampo-neocortical consolidation circuits. FM appears to be most efficient in children, in whom hippocampus and episodic memory are not fully developed (Bauer, 2008). Clinical investigations in patients with MTL lesions have shown that explicit exposure to new information results in poor behavioral outcomes, while FM learning regimes, on the one hand, lead to successful acquisition, and, on the other hand, are hampered by neocortical damage (Sharon et al., 2011; Warren and Duff, 2014). BOLD-fMRI studies in healthy adults show that FM, in comparison to EE tasks, activates a more widespread neocortical network during encoding, which seems to most reliably include the anterior-temporal lobe, ATL (Atir-Sharon et al., 2015; Merhav et al., 2015). Left ATL neocortex, in turn, has been repeatedly suggested as a seat of lexico-semantic representations, playing the role of a central “hub” in distributed word memory circuits (Patterson et al., 2007). Furthermore, while EE seems to benefit from an overnight consolidation stage, learning via fast mapping does not trigger overnight changes in brain representations (Merhav et al., 2015). Moreover, even passive exposure to unattended novel word forms presented repeatedly outside of any task or context leads to immediate changes in the brain responses, indicative of a novel memory trace build-up in the perisylvian neocortex (Kimppa et al., 2015, 2016; Partanen et al., 2017, 2018), provided the exposure is intensive enough (dozens to hundreds of repetitions). Such different brain signatures of the two learning strategies in themselves support (partially) different mechanisms underpinning them and may thus explain diverging learning dynamics and efficiency. In sum, even though in real-life situations the distinctions between the two strategies may be blurred, with both mechanisms at play simultaneously depending on the context and the learning environment, the available evidence allows to conclude that they can be dissociated at the conceptual level as well as behaviorally and neurophysiologically.

However, these findings still leave a number of questions open. First, findings of any advantages offered by FM and/or differential learning outcomes of the two regimes have been questioned by some studies that failed to replicate them (see, e.g., Greve et al., 2014). On the other hand, in spite of frequent claims of FM benefits for learning, most of the above studies in fact show better recognition rates after EE (although this

does not *per se* undermine the distinctions found between the brain mechanisms). Second, the behavioral routines typically used to contrast the learning regimes differ in more than one dimension. The most typical paradigm used to implement this (see e.g., Merhav et al., 2015) uses a word-picture association approach, in which the FM condition presents the subject with two or more images, only one of them being novel, thus requiring inference to understand which of the objects the new word refers to (e.g., “does the glorp have leaves?”); at the same time, the EE condition often presents only a single image in conjunction with its name (“this is a glorp”). Such a design implies a lack of basic visual balancing between the two conditions, which puts differential load already at the level of initial visual processing of the stimuli. Furthermore, at the higher cognitive level, it creates different distribution of attention across the visual field between the two conditions. Whereas attention and executive control can certainly influence learning outcomes (Kimppa et al., 2016), it is important to disentangle their effects and those more directly related to memory or language systems as such.

Third, while these two conditions inevitably frame the task in cognitively different manners, it is further exacerbated by the way the instruction is typically offered in such an experiment. In FM condition (Carey and Bartlett, 1978; Atir-Sharon et al., 2015), a question (“does the glorp have leaves?”) or a request (“bring/show me the glorp”) are used, whereas naming is used in EE (“this is a glorp”). Pragmatically, Naming, Question and Request constitute different speech acts (Searle, 1969) that put different demands on the cognitive system and are known to be underpinned by overlapping yet distinct brain networks (van Ackeren et al., 2012; Egorova et al., 2014), which further confounds any behavioral and neurophysiological distinctions found between FM and EE. While it may not be possible to fully balance the two clearly distinct learning regimes, minimizing the effects of any extraneous factors, such as visual features, attention, cognitive load, and contextual framing, it is highly desirable to disentangle their mechanisms with fewer confounding factors.

More generally, studies diverge hugely in how they train their subjects with new words. This could be word-picture associations that use written or spoken forms or both modalities (Breitenstein et al., 2005), purely sentential context (Mestres-Missé et al., 2007, 2008) or even isolated word forms with no semantics (Gaskell and Dumay, 2003; Shtyrov et al., 2010). Some of the studies use perceptual exposure, while others introduce articulation as an ecological part of the learning process (Rauschecker et al., 2008). Similar to the points above, different learning modalities would introduce the variability into results, complicating the overall picture. Direct comparisons of visual vs. auditory mode of acquisition (the latter being the “native” modality of language), learning in vs. outside context, with vs. without semantic reference, perceptually only vs. with articulation etc. would be important to disentangle all of these factors.

Equally important is the assessment of the learning outcomes. The tasks used for this diverge across studies, and most often include free recall, lexical decision and familiarity judgement. These are more shallow lexical tasks, which may not require full lexico-semantic access of the newly formed memory trace. A more elaborate testing that could require lexical as well as

semantic (e.g., semantic judgement task, semantic matching, free-form definition), and possibly even contextual levels of testing, would therefore be desirable. Further, the assessment of semantics acquisition could also be done on the basis of brain activation patterns, such as the recruitment of meaning-dependant modality-specific networks (Macedonia et al., 2011; Vukovic and Shtyrov, 2014; Mayer et al., 2015).

On a similar note, many studies limit themselves to immediate post-experimental testing, ignoring the longer-term consolidation processes that play a significant role in (at least some types) of acquisition (McMurray et al., 2016). Ideally, the assessment of the learning outcomes should be done both immediately and after an overnight sleep period; longer-term retention of stimulus materials over weeks/months could also be addressed where possible.

Finally, and importantly, the bulk of previous research in this area was done behaviorally and/or using slow neuroimaging tools, such as fMRI, to address distinctions between learning regimes. As such, these measures cannot address rapid neuronal activations that are known to take place on the millisecond range; this is particularly important for language, a function that relies on temporally dynamic processing of information rapidly unfolding over time (Friederici, 2002; Pulvermüller et al., 2009; MacGregor et al., 2012; Shtyrov and Stroganova, 2015). To better understand the neural processes underpinning different types of language learning, there is a need for a more direct measure of electric neuronal activity, which can be provided by time-resolved imaging tools such as EEG or MEG, or, ideally a combination of tools, such as MRI-based source analysis of combined multichannel EEG-MEG data. On the flip-side, while activity patterns obtained in brain studies are useful, causal evidence is also needed to scrutinize these distinctions in healthy individuals. Outside of limited patient studies, such evidence is presently lacking. The use of targeted neurostimulation techniques (such as TMS or tDCS) to influence the learning outcomes may provide the much needed evidence for the involvement of particular brain areas in specific learning types.

On a more conceptual level, the use of learning strategies might differ according to the learning environment, resources, and purposes, while their effectiveness may also vary depending on the learner's age, neural development, cognitive capacities, and overall context. Furthermore, in the natural language acquisition scenario (other than classroom settings), word acquisition, whether in the first or second language, is unlikely to occur exclusively through only one or the other strategy. Instead, both

strategies may be used concurrently which may possibly result in enhanced learning outcomes, although the extent to which each strategy is used depends on the learning environment and the language (first or second) in question. Notably, the brain networks implicated in the two mechanisms do overlap (most importantly in the temporal lobe) and the tight connectivity, which is known to exist between these structures (Catani and Mesulam, 2008; Friederici, 2012), provides for seamless information exchange across the circuits involved. Furthermore, a range of other processes involved in learning (e.g., association, differentiation, enrichment, retrieval) may interactively influence the acquisition of new materials at different stages. Finally, the explicit/implicit distinction is also present in more general models of language, not just word acquisition (e.g., Ullman, 2001; Paradis, 2009). These and similar factors should also be considered in studies investigating learning strategies.

To conclude, the literature to date clearly suggests overlapping yet dissociable learning systems that support different routes of novel word acquisition by the human brain. They diverge in their speed and underlying brain structures, and may be used to different extents for explicitly acquiring presented information or for contextually-driven implicit inference-based learning. The studies available to date diverge in the methodologies employed and present a somewhat controversial picture. To fill these gaps in the field, future studies should use a combination of rigorously matched behavioral regimes, controlled modes of presentation, a comprehensive set of tasks to assess the outcomes at different times, and different neuroimaging tools able to assess both the complex spatio-temporal dynamics of word acquisition, and the causal relationships between brain structures and learning strategies.

AUTHOR CONTRIBUTIONS

YS conceived the original idea for the paper. All authors wrote and revised the manuscript and have approved the content for publication.

ACKNOWLEDGMENTS

Supported by the Lundbeck Foundation (grants R140-2013-12951 and R164-2013-15801, projects 15480, 18690), Danish Council for Independent Research (DFR 6110-00486, project 23776), and RF Government (grant contract No. 14.W03.31.0010).

REFERENCES

- Atir-Sharon, T., Gilboa, A., Hazan, H., Koilis, E., and Manevitz, L. M. (2015). Decoding the formation of new semantics: MVPA investigation of rapid neocortical plasticity during associative encoding through fast mapping. *Neural Plasticity* 2015:804385. doi: 10.1155/2015/804385
- Battaglia, F. P., Benchenane, K., Sirota, A., Pennartz, C. M., and Wiener, S. I. (2011). The hippocampus: hub of brain network communication for memory. *Trends Cogn. Sci.* 15, 310–318. doi: 10.1016/j.tics.2011.05.008
- Bauer, P. J. (2008). Toward a neuro-developmental account of the development of declarative memory. *Dev. Psychobiol.* 50, 19–31. doi: 10.1002/dev.20265
- Bloom, P., and Markson, L. (1998). Capacities underlying world learning. *Trends Cogn. Sci.* 2, 67–73. doi: 10.1016/S1364-6613(98)01121-8
- Breitenstein, C., Jansen, A., Deppe, M., Foerster, A. F., Sommer, J., Wolbers, T., et al. (2005). Hippocampus activity differentiates good from poor learners of a novel lexicon. *NeuroImage* 25, 958–968. doi: 10.1016/j.neuroimage.2004.12.019
- Carey, S., and Bartlett, E. (1978). "Acquiring a single new word," in *Papers and Reports on Child Language Development*, 17–29. Retrieved from: <http://www.>

- mendeley.com/research/acquiring-single-new-word-1/ (accessed February 28, 2019).
- Catani, M., and Mesulam, M. (2008). The arcuate fasciculus and the disconnection theme in language and aphasia: history and current state. *Cortex* 44, 953–961. doi: 10.1016/j.cortex.2008.04.002
- Davis, M. H., and Gaskell, M. G. (2009). A complementary systems account of word learning: neural and behavioural evidence. *Phil. Trans. R. Soc. B.* 364, 3773–3800. doi: 10.1098/rstb.2009.0111
- Dollaghan, C. (1985). Child meets word: “fast mapping” in preschool children. *J. Speech Hear. Res.* 28, 449–454. doi: 10.1044/jshr.2803.454
- Egorova, N., Pulvermüller, F., and Shtyrov, Y. (2014). Neural dynamics of speech act comprehension: an MEG study of naming and requesting. *Brain Topogr.* 27, 375–392. doi: 10.1007/s10548-013-0329-3
- Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends Cogn. Sci.* 6, 78–84. doi: 10.1016/S1364-6613(00)01839-8
- Friederici, A. D. (2012). The cortical language circuit: from auditory perception to sentence comprehension. *Trends Cogn. Sci.* 16, 262–268. doi: 10.1016/j.tics.2012.04.001
- Gaskell, M. G., and Dumay, N. (2003). Lexical competition and the acquisition of novel words. *Cognition* 89, 105–132. doi: 10.1016/S0010-0277(03)00070-2
- Greve, A., Cooper, E., and Henson, R. N. (2014). No evidence that “fast-mapping” benefits novel learning in healthy Older adults. *Neuropsychologia* 60, 52–59. doi: 10.1016/j.neuropsychologia.2014.05.011
- Halberda, J. (2006). Is this a dax which I see before me? Use of the logical argument disjunctive syllogism supports word-learning in children and adults. *Cogn. Psychol.* 53, 310–344. doi: 10.1016/j.cogpsych.2006.04.003
- Horst, J. S., and Samuelson, L. K. (2008). Fast mapping but poor retention by 24-month-old infants. *Infancy* 13, 128–157. doi: 10.1080/15250000701795598
- Kaminski, J., Call, J., and Fischer, J. (2004). Word learning in a domestic dog: evidence for “fast mapping.” *Science* 304, 1682–1683. doi: 10.1126/science.1097859
- Kimppa, L., Kujala, T., Leminen, A., Vainio, M., and Shtyrov, Y. (2015). Rapid and automatic speech-specific learning mechanism in human neocortex. *NeuroImage* 118, 282–291. doi: 10.1016/j.neuroimage.2015.05.098
- Kimppa, L., Kujala, T., and Shtyrov, Y. (2016). Individual language experience modulates rapid formation of cortical memory circuits for novel words. *Sci. Rep.* 6, 1–10. doi: 10.1038/srep30227
- Macedonia, M., Müller, K., and Friederici, A. D. (2011). The impact of iconic gestures on foreign language word learning and its neural substrate. *Hum. Brain Mapp.* 32, 982–998. doi: 10.1002/hbm.21084
- MacGregor, L. J., Pulvermüller, F., Van Casteren, M., and Shtyrov, Y. (2012). Ultra-rapid access to words in the brain. *Nat. Commun.* 3:711. doi: 10.1038/ncomms1715
- Mayer, K. M., Yildiz, I. B., Macedonia, M., and von Kriegstein, K. (2015). Visual and motor cortices differentially support the translation of foreign language words. *Curr. Biol.* 25, 530–535. doi: 10.1016/j.cub.2014.11.068
- McClelland, J. L., McNaughton, B. L., and O’Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychol. Rev.* 102, 419–457. doi: 10.1037/0033-295X.102.3.419
- McMurray, B., Kapnoula, E., and Gaskell, M. G. (2016). “Learning and integration of new wordforms: Consolidation, pruning and the emergence of automaticity,” in *Speech Perception and Spoken Word Recognition*, eds M. G. Gaskell and J. Mirkovic (London: Taylor and Francis), 116–142. doi: 10.4324/9781315772110
- Merhav, M., Karni, A., and Gilboa, A. (2015). Not all declarative memories are created equal: fast mapping as a direct route to cortical declarative representations. *NeuroImage* 117, 80–92. doi: 10.1016/j.neuroimage.2015.05.027
- Mestres-Missé, A., Càmarà, E., Rodriguez-Fornells, A., Rotte, M., and Münte, T. F. (2008). Functional neuroanatomy of meaning acquisition from context. *J. Cogn. Neurosci.* 20, 2153–2166. doi: 10.1162/jocn.2008.20150
- Mestres-Missé, A., Rodriguez-Fornells, A., and Münte, T. F. (2007). Watching the brain during meaning acquisition. *Cerebral Cortex* 17, 1858–1866. doi: 10.1093/cercor/bhl094
- O’Reilly, R. C., and McClelland, J. L. (1994). Hippocampal conjunctive encoding, storage, and recall: avoiding a trade-off. *Hippocampus*. doi: 10.1002/hipo.450040605
- Paradis, M. (2009). *Declarative and Procedural Determinants of Second Languages*. Amsterdam: John Benjamins.
- Partanen, E., Leminen, A., de Paoli, S., Bundgaard, A., Kingo, O. S., Krøjgaard, P., et al. (2017). Flexible, rapid and automatic neocortical word form acquisition mechanism in children as revealed by neuromagnetic brain response dynamics. *NeuroImage* 155, 450–459. doi: 10.1016/j.neuroimage.2017.03.066
- Partanen, E. J., Leminen, A., Cook, C., and Shtyrov, Y. (2018). Formation of neocortical memory circuits for unattended written word forms: neuromagnetic evidence. *Sci. Rep.* 8:15829. doi: 10.1038/s41598-018-34029-y
- Patterson, K., Nestor, P. J., and Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nat. Rev. Neurosci.* 8, 976–987. doi: 10.1038/nrn2277
- Pulvermüller, F., Shtyrov, Y., and Hauk, O. (2009). Understanding in an instant: neurophysiological evidence for mechanistic language circuits in the brain. *Brain Lang.* 110, 81–94. doi: 10.1016/j.bandl.2008.12.001
- Rauschecker, A. M., Pringle, A., and Watkins, K. E. (2008). Changes in neural activity associated with learning to articulate novel auditory pseudowords by covert repetition. *Hum. Brain Mapp.* 29, 1231–1242. doi: 10.1002/hbm.20460
- Scoville, W. B., and Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *J. Neurol. Neurosurg. Psychiatry* 20, 11–21. doi: 10.1136/jnnp.20.1.11
- Searle, J. R. (1969). *Speech Acts: An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.
- Sharon, T., Moscovitch, M., and Gilboa, A. (2011). Rapid neocortical acquisition of long-term arbitrary associations independent of the hippocampus. *Proc. Natl. Acad. Sci.* 108, 1146–1151. doi: 10.1073/pnas.1005238108
- Shtyrov, Y. (2012). Neural bases of rapid word learning. *Neuroscientist* 18, 312–319. doi: 10.1177/1073858411420299
- Shtyrov, Y., Nikulin, V. V., and Pulvermüller, F. (2010). Rapid cortical plasticity underlying novel word learning. *J. Neurosci.* 30, 16864–16867. doi: 10.1523/JNEUROSCI.1376-10.2010
- Shtyrov, Y. Y., and Stroganova, T. A. (2015). When ultrarapid is ultrarapid: on importance of temporal precision in neuroscience of language. *Front. Hum. Neurosci.* 9, 1–5. doi: 10.3389/fnhum.2015.00576
- Suzuki, W. A. (2006). Encoding new episodes and making them stick. *Neuron* 50, 19–21. doi: 10.1016/j.neuron.2006.03.029
- Talpos, J. C., Dias, R., Bussey, T. J., and Saksida, L. M. (2008). Hippocampal lesions in rats impair learning and memory for locations on a touch-sensitive computer screen: the “ASAT” task. *Behav. Brain Res.* 192, 216–225. doi: 10.1016/j.bbr.2008.04.008
- Ullman, M. T. (2001). A neurocognitive perspective on language: the declarative/procedural model. *Nat Rev Neurosci* 2, 717–726. doi: 10.1038/35094573
- van Ackeren, M. J., Casasanto, D., Bekkering, H., Hagoort, P., and Rueschemeyer, S. A. (2012). Pragmatics in action: indirect requests engage theory of mind areas and the cortical motor network. *J. Cogn. Neurosci.* 24, 2237–47. doi: 10.1162/jocn_a_00274
- Vukovic, N., and Shtyrov, Y. (2014). Cortical motor systems are involved in second-language comprehension: evidence from rapid mu-rhythm desynchronisation. *NeuroImage* 102, 695–703. doi: 10.1016/j.neuroimage.2014.08.039
- Walker, M. P., and Stickgold, R. (2006). Sleep, memory, and plasticity. *Ann. Rev. Psychol.* 57, 139–166. doi: 10.1146/annurev.psych.56.091103.070307
- Warren, D. E., and Duff, M. C. (2014). Not so fast: Hippocampal amnesia slows word learning despite successful fast mapping. *Hippocampus* 24, 920–933. doi: 10.1002/hipo.22279

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Shtyrov, Kirsanov and Shcherbakova. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Semantic Processing in Bilingual Aphasia: Evidence of Language Dependency

Marco Calabria^{1*}, Nicholas Grunden^{1,2}, Mariona Serra¹, Carmen García-Sánchez² and Albert Costa^{1,3}

¹ Center for Brain and Cognition, Pompeu Fabra University, Barcelona, Spain, ² Hospital de la Santa Creu i Sant Pau, Barcelona, Spain, ³ Institució Catalana de Recerca i Estudis Avançats, Barcelona, Spain

Individuals with aphasia frequently show lexical retrieval deficits due to increased interference of semantically related competitors, a phenomenon that can be observed in tasks such as naming pictures grouped by semantic category. These deficits are explained in terms of impaired semantic control, a set of abilities that are to some extent dependent upon executive control (EC). However, the extent to which semantic control abilities can be affected in a second and non-dominant language has not been extensively explored. Additionally, findings in healthy individuals are inconclusive regarding the degree to which semantic processing is shared between languages. In this study, we explored the effect of brain damage on semantic processing by comparing the performance of bilingual individuals with aphasia on tasks involving semantic control during word production and comprehension. Furthermore, we explored whether semantic deficits are related to domain-general EC deficits. First, we investigated the naming performance of Catalan–Spanish bilinguals with fluent aphasia and age-matched healthy controls on a semantically blocked cyclic naming task in each of their two languages (Catalan and Spanish). This task measured semantic interference in terms of the difference in naming latencies between pictures grouped by the same semantic category or different categories. Second, we explored whether lexical deficits extend to comprehension by testing participants in a word-picture matching task during a mixed language condition. Third, we used a conflict monitoring task to explore the presence of EC deficits in patients with aphasia. We found two main results. First, in both language tasks, bilingual patients' performances were more affected than those of healthy controls when they performed the task in their non-dominant language. Second, there was a significant correlation between the speed of processing on the EC task and the magnitude of the semantic interference effect exclusively in the non-dominant language. Taken together, these results suggest that lexical retrieval may be selectively impaired in bilinguals within those conditions where semantic competition is higher, i.e.,- in their non-dominant language; this could possibly be explained by an excessive amount of inhibition placed upon this language. Moreover, lexico-semantic impairments seem to be at least somewhat related to conflict monitoring deficits, suggesting a certain degree of overlap between EC and semantic control.

Keywords: bilingual aphasia, semantic control, cycling naming, language dependency, executive control, language control

OPEN ACCESS

Edited by:

Beatriz Martín-Luengo,
National Research University Higher
School of Economics, Russia

Reviewed by:

Yasmeen Farooqi-Shah,
University of Maryland, College Park,
United States

Claudia Peñaloza,

Boston University, United States

*Correspondence:

Marco Calabria
calabria.marc@gmail.com

Received: 28 January 2019

Accepted: 29 May 2019

Published: 14 June 2019

Citation:

Calabria M, Grunden N, Serra M,
García-Sánchez C and Costa A
(2019) Semantic Processing
in Bilingual Aphasia: Evidence
of Language Dependency.
Front. Hum. Neurosci. 13:205.
doi: 10.3389/fnhum.2019.00205

INTRODUCTION

Lexical retrieval deficits in aphasia have many different potential sources of impairment including dysfunction in semantic selection, lexical selection, and/or phonological processing (Laine and Martin, 2006). Some more recent views, specifically those that take into account connectionist models, have broadly defined two main levels of retrieval: the first stage comprised of meaning and grammar and the second of phonological structure and content (Schwartz, 2014). In the present study, we aim to investigate the role of semantic control, defined as one of the mechanisms within the semantic network, in lexical retrieval deficits within patients with bilingual aphasia.

Semantic control can be defined as a set of processes that enable an individual to modulate retrieval of information based on the contextual cognitive demand (Lambon Ralph et al., 2017) and can be distinguished from semantic representation within the semantic cognition network (Jefferies, 2013; Lambon Ralph et al., 2017). To some extent, this idea coincides with the concept of ‘access deficits’ in semantic aphasia as opposed to the ‘storage deficits’ in semantic dementia (for a review, see Mirman and Britt, 2004). Of particular interest within the context of post-stroke aphasia is the control element of this semantic framework, since semantic memory is usually spared (Jefferies and Lambon Ralph, 2006; Jefferies et al., 2008; Rogers et al., 2015).

Findings from neuroimaging studies have distinguished a semantic control network that includes the left prefrontal and temporo-parietal cortices, as opposed to the anterior temporal lobes, serving as crucial elements for conceptual representations (Noonan et al., 2010, 2013). Interestingly, the neural model of semantic cognition proposed by Lambon Ralph et al. (2017) shows an overlap with the bilingual language control network in which prefrontal areas are engaged in conflict resolution and the posterior areas (inferior parietal lobules) in language selection (Abutalebi and Green, 2016; Calabria et al., 2018). Most studies agree that a second language (L2) is mainly acquired through the same neural devices responsible for the first language (L1) and that the brain systems associated with the linguistic processing are shared between the two languages (e.g., Perani and Abutalebi, 2005; Abutalebi and Green, 2007). Broadly speaking, we did not expect there to be a difference in semantic control abilities for L1 and L2. However, some differences have been reported between the two languages when bilinguals have to process semantic incongruence. In their review, Moreno et al. (2008) concluded that semantic processing in L2 is delayed, as measured by a delayed peak latency of the event-related potential (N400) associated with semantic violation, thus suggesting differences in semantic integration between the bilinguals’ two languages. Similarly, some bilingual models of speech production claimed that lexico-semantic representation might function differently for a bilingual’s two languages (Kroll and Stewart, 1994; Gollan et al., 2008; Kroll et al., 2015; for a review see Branzi et al., 2018).

In the present study, we wanted to test the hypothesis of language-independency of semantic control by investigating the performance of bilingual patients with aphasia on semantic control tasks in their two languages (Catalan and Spanish). To do so, we used the semantic blocked cycling naming task that has

been extensively used to investigate semantic interference both in healthy individuals (Damian and Bowers, 2003; Belke et al., 2005; Damian and Als, 2005; Navarrete et al., 2012; Belke, 2017) and in monolingual patients with aphasia (McCarthy and Kartsounis, 2000; Wilshire and McCarthy, 2002; Schnur et al., 2006; Biegler et al., 2008; Harvey and Schnur, 2015) as a measurement of semantic competition during lexical selection.

Semantic Processing in Healthy Bilinguals: Language-Dependent or Independent?

The results of a series of behavioral and neuroimaging studies agree with the hypothesis that there are similar principles of semantic representation across languages. For instance, studies that used semantic cross-language priming found that, with highly proficient bilinguals, the magnitude of word priming between semantically related words is similar irrespective of the language direction (e.g., Zeelenberg and Pecher, 2003; Perea et al., 2008; Travis et al., 2017). Furthermore, when bilinguals have to name pictures in a semantically demanding task, they show a similar magnitude of semantic interference in both L1 and L2, suggesting that semantic control abilities are independent of the language being utilized (Runnqvist et al., 2012).

Despite that some qualitative differences between languages have been found, the main results of relevant studies support the hypothesis of a shared conceptual/semantic system across languages (Francis, 1999, 2005), as proposed in some models of bilingual production and comprehension (BIA+ model: Dijkstra and van Heuven, 2003; ICM: Green, 1986; RHM: Kroll and Stewart, 1994).

A second line of research has investigated the underlying neural network of semantic processing in bilinguals on a variety of semantic tasks. Some studies concur that bilinguals show similar activation while they are processing semantic representations in their L1 and L2, identifying a language-invariant semantic network that includes the inferior temporal lobe (Grogan et al., 2009), superior temporal lobes (Chee et al., 2001; Pillai et al., 2003), frontal (Illes et al., 1999; Chee et al., 2001) or a more widely distributed set of language areas (Correia et al., 2014; Van de Putte et al., 2017). One exception is a study conducted by Klein et al. (2006) that found activation in the putamen when subjects performed L1-L2 translation but not the inverse direction that coincided with an otherwise complete overlap of activation for the two languages during a word generation task.

Finally, some evidence of the possible language-dependency of semantic processing comes from sentence processing in bilinguals. Specifically, the event-related potential component (N400) that indexes semantic violation has been found to be consistently delayed in its peak latency for L2 relative to L1 (for a review see Moreno et al., 2008).

Therefore, although most studies agree that bilinguals show shared semantic networks for L1 and L2, some research revealed the presence of language-dependent processes, possibly related to the type of task used to assess semantic representation or control. These results are in line with what some other models

of bilingual speech production have proposed. For instance, the Revised Hierarchical Model (RHM) by Kroll and Stewart (1994) assumes that the L1 lexicon is larger than that of L2 and that the connections between L1 concepts are stronger than those between L2, which are thought to be attached to the L1 lexicon. Similarly, the ICM model by Green (1986) would predict different degrees of inhibitory control in each language that, once applied at the schema level, would modulate lexical selection according to the dominance of the two languages.

Semantic Deficits in Bilingual Speakers With Brain Damage

Research that has investigated semantic deficits in bilingual patients with neurodegenerative disorders has shown similar impairments across languages, suggesting that semantic processing is language-independent (Mendez et al., 1999; Hernández et al., 2008, 2010). In the first study by Hernández et al. (2008), patient JPG had similar category-specific deficits in both languages (Spanish and Catalan) with worse performance in naming verbs than nouns. In a further study, Hernández et al. (2010) found that the semantic memory deficits of JFF (a Catalan–Spanish bilingual patient) had a similar influence on his performance while he performed word translation in both language directions. In both studies, only some qualitative differences of errors between languages were reported, but the main result supports a shared conceptual representation across languages (Francis, 1999, 2005).

Also, studies performed on bilingual patients with aphasia have found that the representational level of knowledge is shared between languages. For example, Siyambalapitiya et al. (2013) found that their patient (SN) not only showed intact semantic priming in both languages, but also in the cross-language condition (from English to Spanish), supporting a language-independent nature of bilingual semantic memory. Other research within bilingual aphasia has uncovered a more complex picture that would support the notion that, in post-stroke aphasia, patients' deficits arose from dysfunction in the control part of the semantic system instead of representational system of knowledge as in dementia patients [see the controlled semantic cognition (CSC) model by Lambon Ralph et al., 2017]. Some of this data comes from the study of cross-language generalization using semantic-based training (for a more extended discussion on cross-language issues in aphasia see Faroqi-Shah et al., 2010; Khachatryan et al., 2016). For instance, Kiran and Roberts (2010) found that the only one of the four patients they tested after semantic-based naming treatment improved in the untrained language, suggesting that providing semantic information to improve lexical retrieval has little to no cross-language transfer. Kiran et al. (2013b) in a further study found a similar result of limited cross-language generalization for semantic representations. Interestingly, they proposed that the degree of cross-language transfer might be explained by the integrity of two independent mechanisms: the first being a generalized mechanism involved in the spreading of activation brought about via treatment and the second being inhibitory control which, in the case of bilingual speakers, would

interfere with the activation level of their two languages (Green, 1986). Therefore, the degree of within- and between-language generalization depends on the interplay of these two mechanisms, namely how inhibitory control works to allow semantic activation to increase in one language and/or in both.

Interestingly, the idea that EC plays a role during semantic processing is similar to what was proposed by Lambon Ralph et al. (2017) for monolinguals. These authors claim that, along with an amodal 'hub' which functions by integrating different sources of information (Patterson et al., 2007), there are EC mechanisms that supervise how activation spreads throughout the semantic representation network. That is, there exists a combination of two systems: one representational (temporo-parietal) and one for control (frontally distributed), with the latter being more closely related to semantic control deficits in monolingual patients with aphasia (Harvey and Schnur, 2015).

Therefore, following the idea of the CSC model, we aimed to investigate whether semantic control may be differentially affected in the two languages of bilinguals post-stroke. To do so, we employed a blocked naming task that allowed us to manipulate the amount of interference during word retrieval for semantically-related competitors. This type of paradigm has been extensively used in studies with monolingual patients with aphasia to test the root causes of word retrieval deficits (e.g., Wilshire and McCarthy, 2002; Biegler et al., 2008; Schnur et al., 2009; Scott and Wilshire, 2010; Harvey and Schnur, 2015). According to some authors, this task can also help test whether word retrieval deficits can be explained in terms of an increased excitation or an excessive inhibition applied to semantic competitors, resulting in the target words being less available during lexical selection (Schnur et al., 2006).

Moreover, to specifically test the relationship between semantic control processes and EC, we tested patients on a conflict monitoring task. The inclusion of this task was motivated by a new body of research with bilingual aphasic patients that highlights the cross-talk between deficits in language control and in domain-general EC (Dash and Kar, 2014; Gray and Kiran, 2015; Faroqi-Shah et al., 2016; for a review on this issue see Calabria et al., 2018).

The Present Study

To investigate semantic control during speech production in patients, we employed a semantic blocked cyclic naming task. In this paradigm, participants were required to name pictures in two conditions: (a) homogeneous, where pictures belonged to the same semantic category (e.g., only animals), and (b) heterogeneous, where pictures belonged to different semantic categories (e.g., animals, furniture, tools, etc.). The latencies in the naming of elements in the heterogeneous condition become faster over repetitions (cycles) whereas those in the homogeneous generally remain constant after the second cycle (e.g., Damian and Bowers, 2003; Belke et al., 2005; Damian and Als, 2005; Navarrete et al., 2012; Crowther and Martin, 2014; Belke, 2017). The difference in naming latencies between these two conditions is an index of semantic interference that is increased in patients with aphasia compared to healthy individuals (e.g., Schnur et al., 2006, 2009; Biegler et al., 2008; Scott and Wilshire, 2010)

due to hyper-activation or excessive inhibition of semantic competitors brought on by their language impairments. This agrees with the view that lexical selection is a competitive process (for a recent review see Nozari and Hepner, 2018). The automatic activation of semantically related items spreads to their corresponding lexical representations and the target word competes for selection (for non-competitive models see Costa et al., 1999; Mahon et al., 2007).

The general hypothesis about semantic control in bilingualism was that if semantic control was language-independent, we expected to see a similar increase of semantic interference in both languages in patients with aphasia compared to healthy controls. Indeed, according to the models that have proposed that lexical selection in bilinguals is qualitatively similar to that of monolinguals, we should expect language-independency of semantic control (Costa et al., 1999; Caramazza and Costa, 2000; La Heij, 2005; Finkbeiner et al., 2006).

On the other hand, if semantic control was language-dependent, we expected to see higher interference in one language compared to the other. Presumably, more semantic impairment would occur in the non-dominant language if it were related to EC deficits (e.g., Abutalebi and Green, 2007, 2016) or had weaker connections between lexical and semantic units (e.g., Kroll and Stewart, 1994).

In order to assess the integrity of semantic representations, we employed a bilingual word-picture matching task. Participants were required to match a picture presented on the screen with one of two word options (semantically related, same category). The main reasoning behind the inclusion of this task was to measure the accuracy of patients as compared to healthy controls on the task and thus to exclude the possibility of any representational deficits in semantic memory. We adopted a bilingual version of the matching task because this type of paradigm allowed us to test both languages in the same task and because we have already seen previous evidence that it serves as a robust task for testing comprehension in bilinguals (Macizo et al., 2010; Mosca and de Bot, 2017).

Additionally, we investigated the integrity of EC with a conflict monitoring task in patients and healthy controls. This task has been used previously in studies with bilingual patients with the aim to investigate the relationship between language control and EC deficits (Green et al., 2010; Gray and Kiran, 2015). We correlated patients' performance on this EC task with the semantic blocked cyclic naming task, with the degree of the correlation indicating to what extent the two domains of control overlap. The available literature on this issue reports mixed findings and the number of studies performed with bilingual patients after stroke in which the two domains have been compared is very slim (for a review see Calabria et al., 2018), resulting in a need for further research on this issue. Therefore, an overlap of deficits in both domains would suggest that domain-general EC is also involved in language selection. That is, the hyper-activation or -inhibition upon the semantic competitor during lexical selection would be intimately related to non-linguistic EC processes (inhibitory control and/or conflict resolution), as predicted by the 'executive selection account' (Wilshire and McCarthy, 2002).

To summarize, the current study was undertaken to explore two issues in the context of semantic control and bilingualism:

- (1) Are semantic control processes language-dependent and differentially affected by brain damage in a bilingual's two languages?
- (2) Are semantic control processes dependent on general-domain EC mechanisms such as conflict monitoring and resolution?

MATERIALS AND METHODS

Participants

A total of 11 Catalan-Spanish patients with bilingual aphasia were recruited from the Speech Therapy Unit of Hospital de la Santa Creu i Sant Pau in Barcelona. All patients were speakers of both Catalan and Spanish prior to stroke, exhibited adequate hearing and vision, demonstrated stable health status and were in the chronic stage for language disorders (more than 1 year post-injury). The etiology was brain tumor for one patient (Pt2) and cerebrovascular (either ischemic or hemorrhagic stroke) for all other patients. All patients had lesions localized in the left hemisphere.

A group of 13 healthy individuals also participated in the study as controls; their demographic and linguistic characteristics were matched to those of patients with aphasia.

Language Assessment

To define the type and the degree of language impairment, the Spanish version of the Western Aphasia Battery (Kertesz and Pascual-Leone García, 1990) was administered by Dr. García Sánchez, a clinical neuropsychologist with expertise in aphasia from the same hospital. The WAB is a comprehensive test of language functions with a relatively short test administration time (30–60 min) and includes four language subtests which assess spontaneous speech, comprehension, repetition, and naming to calculate an Aphasia Quotient (AQ). Patients were only tested in Spanish since a Catalan version of the WAB is not currently available.

According to WAB assessment, one patient was classified as having conduction aphasia, two with Wernicke's aphasia and eight classified as presenting anomia. The degree of language impairment ranged from mild to moderate (55.6 to 84.5 out of 100) and the mean values for each subtest were: 14.1/20 (± 2.6) for Fluency, 8.2/10 (± 1.2) for Comprehension; 7.4/10 (± 1.7) for Repetition, and 7.4/10 (± 1.1) for Naming (see **Table 1** for details).

Patients' language abilities were also tested using part C of the Bilingual Aphasia Test (BAT, Paradis and Libben, 1987) which assesses cross-language abilities over four subtests: Word Recognition, Word Translation, Sentence Translation, and Grammatical Judgment. In Word Recognition, patients were asked to select the correct translation for each word from a list of 10 possible choices (5 words per language; max. score = 10). In the Word Translation task, patients needed to verbally supply the translation of a word spoken by the examiner (10 words

TABLE 1 | Individual scores of patients for the Western Aphasia Battery and Bilingual Aphasia Test (Part C).

Patients	Aphasia type	Months post-onset	Aphasia quotient* (max. 100)	Severity	Fluency (max. 20)	Comprehension (max. 10)	Repetition (max. 10)	Naming (max. 10)	DL	NDL	BAT-C DL (max. 48)		BAT-C NDL (max. 48)		BAT-C- DL vs. NDL p-values		LogType/ LogToken	
											DL	NDL	DL	NDL	DL	NDL	DL	NDL
P11	ANOMIC	73	71.00	Moderate	12.00	8.50	8.10	8.10	CAST	CAT	29	24	0.405	0.92	0.92	0.92	0.92	0.92
P12	CONDUCTION	131	84.50	Mild	18.00	9.25	6.40	8.60	CAT	CAT	24	19	0.407	0.87	0.87	0.88	0.88	0.88
P13	ANOMIC	122	64.00	Moderate	10.00	7.30	8.20	6.50	CAT	CAT	21	18	0.675	0.85	0.85	0.88	0.88	0.88
P14	ANOMIC	46	73.40	Moderate	15.00	8.40	7.00	6.30	CAT	CAT	29	26	0.675	0.86	0.86	0.87	0.87	0.87
P15	WERNICKE	108	75.70	Mild to moderate	16.00	6.75	6.20	8.90	CAT	CAT	28	24	0.533	0.83	0.83	0.89	0.89	0.89
P16	ANOMIC	49	79.90	Mild to moderate	14.00	9.55	8.60	7.30	CAST	CAT	14	12	0.817	0.91	0.91	0.91	0.91	0.91
P17	ANOMIC	21	88.20	Mild	15.00	9.50	9.70	8.90	CAT	CAT	30	30	0.830	0.86	0.86	0.83	0.83	0.83
P18	ANOMIC	119	57.60	Moderate	10.00	6.90	7.10	5.80	CAT	CAT	16	17	0.828	0.80	0.80	0.80	0.80	0.80
P19	WERNICKE	84	56.30	Moderate	13.00	6.25	3.20	5.70	CAT	CAT	17	9	0.106	0.79	0.79	0.77	0.77	0.77
P110	ANOMIC	66	92.40	Mild	18.00	10.00	9.60	8.10	CAST	CAT	46	36	0.007	0.95	0.95	0.89	0.89	0.89
P111	ANOMIC	14	72.20	Moderate	14.00	7.90	7.30	6.90	CAT	CAST	26	28	0.834	0.87	0.87	0.88	0.88	0.88

Aphasia Quotient (AQ). Range of severity of the scores: 0–25: Very severe; 26–50: Severe; 51–75: Moderate; > 76: Mild. DL: dominant language; NDL: non-dominant language.

per language; max. score = 20). Increasing in difficulty, subjects then were asked in the Sentence Translation task to provide a translation of a sentence that could be repeated a maximum of three times by the examiner (scoring based on correct translations of 3 sections of each sentence for 6 sentences in each language; max. score = 36). Finally, in Grammatical Judgment, patients were asked to determine whether a sentence spoken by the examiner was grammatically correct and, if incorrect, how to fix it (scoring based on correct judgment of grammatical structure and accurate correction of grammatical mistakes if applicable for 8 sentences per language; max. score = 28). These subtests of the BAT-C were administered by a bilingual neuropsychologist, completing all four tasks in one direction of translation followed by the same four tasks in the other direction (i.e., Catalan to Spanish in all tasks followed by Spanish to Catalan).

Furthermore, to have an additional measure of language impairment in their two languages, we asked patients to describe two complex picture scenes: the Cookie Theft Picture (Goodglass and Kaplan, 1972) and the scene description from the WAB. They were instructed to use Catalan to describe the scene in one session and Spanish in the other, with this order counterbalanced across subjects. If some features of the pictures were neglected, the experimenter pointed to them and asked the patient to mention them. Speech was recorded and subsequently analyzed off-line. We collected one recording for each language, each lasting 3 min. After transcribing the descriptions in each language, the total number of words (tokens) and the number of different words (type) were calculated. In order to reduce the impact of sample size, we calculated the individual token-type ratio for each language by using the following transformation $\log_{\text{type}}/\log_{\text{token}}$ (Kong, 2016).

Language Profile

Language history and dominance were determined by means of a questionnaire administered to the participants and an interview with them. Pre-morbid language proficiency in the two languages (Catalan and Spanish) was self-rated by each participant on a four-point scale of their abilities of speaking, comprehension, writing and reading (1 = poor, 2 = regular, 3 = good, 4 = perfect). As can be appreciated in **Table 2**, both patients and healthy controls were highly proficient in all four linguistic domains (see also **Appendix I** for individual data). Moreover, participants were considered early bilinguals as, on average, they were first regularly exposed to their non-dominant language at 6 years of age, thus not differing significantly from the exposure to their dominant language. Finally, language usage was rated based on ten questions in which participants were required to report with what frequency they spoke each of the two languages across different periods of their lives. The final score was transformed into a percentage (from 0 meaning using only Spanish to 100% meaning using only Catalan, around 50% translating to balanced use of the two languages). Both patients and healthy controls reported equal amounts of Catalan and Spanish usage and thus would be considered balanced bilinguals.

The bilinguals that participated in this study acquired their two languages at the same time and it is difficult to say which

TABLE 2 | Socio-demographic and linguistic characteristics of the samples.

	Patients with aphasia		Healthy controls		<i>p</i> -values
	<i>(n = 11)</i>		<i>(n = 13)</i>		
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
Age (years)	58.2	6.4	55.4	4.1	0.29
Education (years)	13.6	1.7	14.4	1.2	0.18
Age of regular dominant language use	2.1	0.7	2.2	0.3	0.52
Age of regular non-dominant language use	4.6	1.6	4.8	1.2	0.75
Language proficiency (1–4)					
<i>Dominant language</i>					
Speaking	4.0	0.0	3.9	0.3	0.35
Comprehension	4.0	0.0	4.0	0.0	–
Reading	3.7	0.6	4.0	0.0	0.21
Writing	3.7	0.6	3.9	0.3	0.21
<i>Non-dominant language</i>					
Speaking	3.9	0.3	4.0	0.0	0.30
Comprehension	3.9	0.3	4.0	0.0	0.30
Reading	3.9	0.4	3.8	0.4	0.51
Writing	3.7	0.6	4.0	0.0	0.16
% Language use	54.5	15.3	45.1	20.8	0.22

would be their L1 or L2. Therefore, we used the terms ‘dominant’ and ‘non-dominant’ instead of L1 and L2 to refer to their languages. The use of ‘dominant’ refers to the language that they prefer to use (or they feel more comfortable speaking), even if they reported that their ‘non-dominant’ language was at the same level of proficiency and frequency of usage as their dominant. According to this definition, 3 patients and 3 healthy controls were classified as Spanish-dominant bilinguals while the rest were classified as Catalan-dominant bilinguals.

Materials and Procedure

The experimental software used for the administration of all tasks was DMDX (Forster and Forster, 2003). All the participants performed three experimental tasks: the semantic blocked cyclic naming task, the bilingual word-picture matching task, and the flanker task. Before starting the experimental procedure, the patients signed an informed consent approved by the ‘Parc de Salut MAR’ Research Ethics Committee under the reference number: 2018/8029/I.

Semantic Blocked Cyclic Naming Task

Stimuli consisted of 32 pictures total with 8 exemplars from 4 semantic categories (animals, vegetables, kitchen tools, and furniture) selected from the Snodgrass and Vanderwart (1980) database (see **Appendix II** for the details of the stimulus). Participants were required to name 8 blocks of pictures: 4 blocks containing semantically related items (Homogenous) and 4 blocks containing semantically unrelated items (Heterogeneous). For some participants, two Homogenous blocks were followed by four Heterogeneous and then two Homogenous blocks

whereas, for others, this pattern was reversed. Sets of 16 different pictures for each language were presented four times (cycles) in 4 Homogenous as well as 4 Heterogeneous blocks, with a total number of 128 naming trials per participant. Eight different lists consisting of 128 stimuli each were created for each language, avoiding the repetition of the same set of pictures between languages.

Each trial included the following elements: a fixation point presented for 750 ms followed by the picture to be named which appeared for up to 2000 ms or until response was provided. After each block, participants were allowed to rest. In order to reduce the number of errors due to possible name disagreement/confusion, participants were presented with the set of pictures before the experimental task and were asked to name them in the required language. Participants were tested in two languages (Catalan and Spanish) and, when possible, over two different sessions staggered week apart. The order of language testing was counterbalanced across participants.

The dependent variables were naming latencies (RTs), which were analyzed off-line with Checkvocal (Protopapas, 2007), and accuracy. Errors were classified as follows: ‘No response,’ when the patient was unable to name the object; ‘semantic,’ when they produced an incorrect word semantically related to the target; ‘cross-language intrusion,’ when they produced the correct word but in the incorrect language; ‘phonological paraphasia,’ when they deleted, substituted or added phonemes to the correct word describing the picture; and ‘unrelated,’ when they produced a word with no relation, semantic or otherwise, to the target word.

Bilingual Word-Picture Matching Task

Stimuli were made up of 60 pictures from different semantic categories selected from Snodgrass and Vanderwart (1980) database. A list of 240 words was also selected consisting of two types of stimuli: (a) 120 as target words corresponding to the picture presented (60 in Catalan and 60 in Spanish); (b) 120 as distractor words semantically related to target words (60 in Catalan and 60 in Spanish). Distractor and target words were of the same semantic category. Each picture was presented with a pair of words, one being the target and the other being the distractor. The pictures and the words were presented in a mixed language condition (Catalan and Spanish), but within each trial the two words were from the same language. There were two types of trials: repeat trials in which participants had to match the picture to a target word in the same language as the target of the previous trial, and switch trials in which participants were required to do the matching within the opposite language compared to the previous trial. There were a total of 120 trials presented in the following manner: 43 Spanish repeat trials, 43 Catalan repeat trials, 17 Spanish switch trials and 17 Catalan switch trials; the task was thus comprised of 28% switch trials and 72% repeat trials. Every trial started with a fixation point (a black cross) in the center of the screen displayed for 500 ms, followed by a picture and two words below for a maximum of 2500 ms. Participants were required to match the target word with the picture presented on the screen by pressing one of two keys on the keyboard. The two keys used for the response

corresponded to the word appearing on the same side of the computer as the key (i.e., “z” corresponding to the word on the left side of the screen). Dependent variables were defined as RTs and accuracy.

Flanker Task

Target stimuli consisted of a row of five horizontal black lines with arrowheads pointing left or right, with the central arrow acting as the true target. Participants were instructed to indicate the direction (left or right) of the central arrow via pressing one of two keys on the keyboard. The target (central arrow) was presented in two main conditions: with congruent flankers (same direction as the target) and incongruent flankers (opposite direction). The event presentation was as follows: (a) a fixation point (a plus sign) appeared at the center of the screen for 400 ms, and (b) the target arrow and the flankers were presented simultaneously until the participants responded or for up to 2000 ms. The experiment consisted of two blocks of 48 trials each, for a total of 96 trials. The proportion of congruent trials was 75% ($n = 72$) to 25% for incongruent trials ($n = 24$). Participants gave their responses by pressing either the ‘V’ or ‘M’ key according to the direction in which the arrow target was pointing. The dependent variables were RTs and accuracy.

RESULTS

Language Impairment in Two Languages

For each participant, we compared the scores of the BAT-C of the two languages using a Chi-squared test with Yates’ correction; ten out of eleven patients showed parallel language deficits (only Pt10 showed a significantly more impaired score in their non-dominant compared to their dominant language, see **Table 1**).

For connected speech, paired t -tests were used to analyze differences between languages (dominant vs. non-dominant); and no difference was found between the two languages in any patient [log type/log token: dominant language = 0.87, non-dominant language = 0.86; $t(10) = -0.09$, $p = 0.92$] (see **Table 1**).

These two results show that our patients had parallel language impairments.

Semantic Blocked Cyclic Naming Task

We first explored the effects of semantic blocking in healthy individuals by performing repeated-measures ANOVAs including Condition (Homogenous vs. Heterogeneous), Language (Dominant vs. Non-dominant), and Cycle (1, 2, 3, and 4) as within-subject factors in the control group only. In a further analysis, we performed repeated-measures ANOVAs including the same within-subject factors and Group as a between-subject factor (patients with aphasia vs. healthy controls). The analyses were performed for two dependent variables—RTs and accuracy—separately. RTs were analyzed for correct responses only. Moreover, RTs across all conditions exceeding three standard deviations above or below mean were excluded from the analyses for each participant.

Reaction Times (RTs)

The analysis with healthy controls showed that main effects of Condition [$F(1,12) = 1307$, $p = 0.004$, $\eta_p^2 = 0.52$] and Cycle [$F(3,36) = 17.41$, $p < 0.001$, $\eta_p^2 = 0.59$] were significant, but not Language [$F(1,12) = 0.05$, $p = 0.82$]. The interaction between Condition and Cycle was also significant [$F(3,36) = 5.79$, $p = 0.002$, $\eta_p^2 = 0.33$]. *Post hoc* analyses showed that, in the Heterogeneous condition, naming latencies became faster over cycle (1st cycle: $M = 712.91$ ms, $SD = 34.88$; 2nd cycle: $M = 664.67$, $SD = 27.47$ ms; 3rd cycle: $M = 639.18$ ms, $SD = 27.01$; 4th cycle: $M = 629.23$ ms, $SD = 24.06$; $p_s < 0.05$). On the other hand, naming latencies in Homogeneous conditions only decreased from the first ($M = 709.32$ ms, $SD = 27.71$) to the second cycle ($M = 672.24$ ms, $SD = 28.72$) ($p = 0.04$). No other interaction was significant.

The analysis that included both groups showed that the main effects of Condition [$F(1,22) = 58.12$, $p < 0.001$, $\eta_p^2 = 0.72$] and Cycle [$F(1,22) = 9.28$, $p < 0.001$, $\eta_p^2 = 0.29$] were significant, but not Language [$F(1,23) = 0.52$, $p = 0.48$]. Also, the main effect of Group was significant [$F(1,22) = 39.79$, $p < 0.001$, $\eta_p^2 = 0.64$] indicating that patients overall were slower ($M = 1107.41$ ms, $SD = 44.05$) than controls ($M = 671.75$ ms, $SD = 46.87$) in performing the task (see **Figure 1** and **Table 3**).

The interaction between Condition and Cycle was also significant [$F(3,66) = 5.25$, $p = 0.003$, $\eta_p^2 = 0.19$]. *Post hoc* analyses showed that in the Heterogeneous condition naming latencies became faster from the first cycle ($M = 960.27$ ms, $SD = 33.98$) to the second ($M = 917.93$ ms, $SD = 34.17$) to the third cycle ($M = 873.25$ ms, $SD = 39.12$) ($p_s < 0.05$). On the other hand, naming latencies in Homogeneous conditions only decreased from the first ($M = 966.16$, $SD = 32.45$ ms) to the second cycle ($M = 918.51$ ms, $SD = 36.78$) ($p = 0.03$).

Finally, the Language \times Condition \times Cycle interaction [$F(3,66) = 4.05$, $p = 0.01$, $\eta_p^2 = 0.15$] as well as the Language \times Condition \times Cycle \times Group interaction were significant [$F(3,69) = 3.15$, $p = 0.03$, $\eta_p^2 = 0.12$]. Further analyses were conducted by comparing the semantic interference effects (difference in naming latencies between the Homogenous and the Heterogeneous condition) within the two groups of participants for each language separately. In the non-dominant language, the magnitude of the semantic interference effect was larger in patients than in controls for the cycles 3 [patients: $M = 178.27$ ms, $SD = 41.90$; controls: $M = 39.69$ ms, $SD = 38.54$; $F(1,24) = 5.92$, $p = 0.02$, $\eta_p^2 = 0.21$] and 4 [patients: $M = 182.36$ ms, $SD = 52.61$; controls: $M = 39.46$ ms, $SD = 48.40$; $F(1,24) = 3.99$, $p = 0.05$, $\eta_p^2 = 0.15$]. In the dominant language, the magnitude of the semantic interference effect did not differ between patients and healthy controls across cycles (all $p_s > 0.05$).

Accuracy

The analysis with healthy controls revealed no main effect or interaction that was statistically significant.

The analysis with both groups showed that the main effect of Group was significant [$F(1,22) = 14.51$, $p = 0.001$, $\eta_p^2 = 0.40$], indicating that the patients’ performance ($M = 82.25\%$, $SD = 3.22$) was lower than that of controls ($M = 98.83\%$, $SD = 2.95$). Also, the

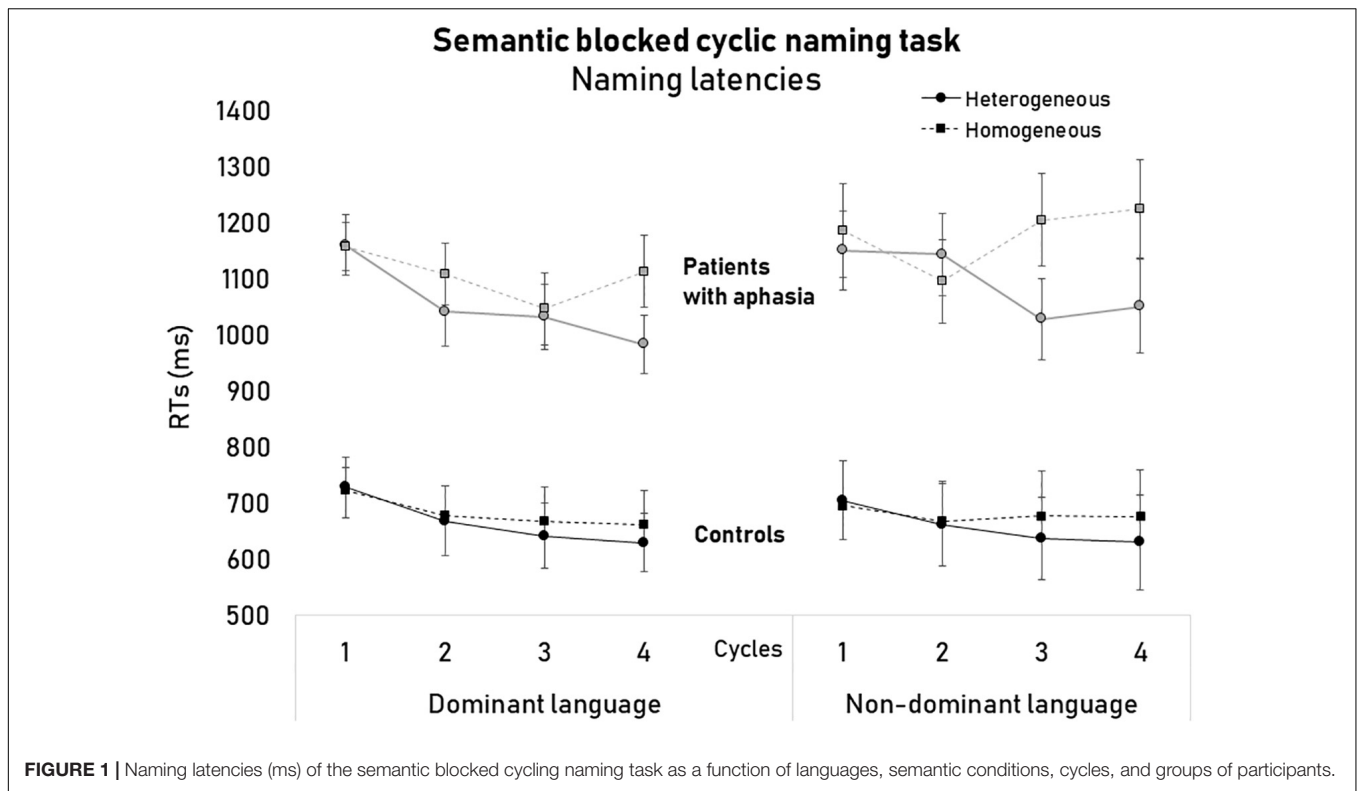


FIGURE 1 | Naming latencies (ms) of the semantic blocked cycling naming task as a function of languages, semantic conditions, cycles, and groups of participants.

main effect of Cycle [$F(3,66) = 7.33, p < 0.001, \eta_p^2 = 0.25$] and the interaction between Cycle and Group [$F(3,66) = 5.61, p = 0.002, \eta_p^2 = 0.20$] were significant. *Post hoc* analyses reveal that patients, but not controls, showed little increase of accuracy in the cycle 3 ($M = 83.93\%, SD = 3.09, p = 0.03$) and 4 ($M = 84.34\%, SD = 2.96, p = 0.02$) compared to the first ($M = 78.91\%, SD = 3.62$).

Error Analysis

The frequency of error types for the two languages is detailed below:

- No response: 54.23% of dominant language errors and 51.10% of non-dominant language errors;
- Phonological errors: 12.31% of dominant language errors and 11.16% of non-dominant language errors;
- Semantic errors: 14.46% of dominant language errors and 15.61% of non-dominant language errors;
- Cross-language intrusions: 9.23% of dominant language errors and 12.26% of non-dominant language errors;
- Unrelated: 9.77% of dominant language errors and 9.87% of non-dominant language errors.

Bilingual Word-Picture Matching Task

In an initial analysis, repeated-measures ANOVAs were performed including Type of Trial (repeat vs. switch) and Language (Dominant vs. Non-dominant) as within-subject factors in healthy controls only. Following said analysis, repeated-measures ANOVAs were performed with the same within-subject factors but also including Group (patients with aphasia vs. healthy controls) as a between-subject factor.

The analyses were performed for two dependent variables—RTs and accuracy—separately. Two patients did not complete this task; therefore, the group comparison was carried out between 10 patients and 13 healthy controls. RTs were analyzed for correct responses only. Moreover, RTs across all conditions exceeding three standard deviations above or below mean were excluded from the analyses for each participant.

TABLE 3 | RTs in the semantically blocked cyclic naming task for healthy controls and patients with aphasia.

Dominant language	Cycle 1		Cycle 2		Cycle 3		Cycle 4	
	M	SE	M	SE	M	SE	M	SE
	Patients with aphasia							
Heterogeneous	1160	54	1041	62	1032	58	983	52
Homogeneous	1158	43	1109	54	1046	64	1113	64
Healthy controls								
Heterogeneous	728	52	668	60	642	55	630	50
Homogeneous	723	41	678	52	668	62	662	61
Non-dominant language								
Patients with aphasia								
Heterogeneous	1151	70	1143	74	1028	73	1051	84
Homogeneous	1186	84	1095	74	1205	83	1225	87
Healthy controls								
Heterogeneous	705	68	661	71	637	70	630	81
Homogeneous	695	80	668	71	678	80	676	84

Reaction Times

The analysis with healthy controls revealed no main effect or interaction that was statistically significant [Type of Trial: $F(1,12) = 0.87, p = 0.37$; Language: $F(1,12) = 0.23, p = 0.64$; Type of Trial \times Language: $F(1,12) = 0.03, p = 0.86$].

In the analysis with both groups, the main effect of Type of Trial [$F(1,22) = 3.28, p = 0.08$] and Language [$F(1,22) = 1.57, p = 0.22$] were not statistically significant. However, the main effect of Group was significant [$F(1,22) = 57.85, p < 0.001, \eta_p^2 = 0.72$], indicating that patients ($M = 1942.52$ ms, $SD = 75.45$) performed more slowly than healthy controls ($M = 1051.88$ ms, $SD = 72.24$). Also, the interactions between Type of Trial and Language [$F(1,22) = 4.95, p = 0.04, \eta_p^2 = 0.18$] and Type of trial \times Language \times Group [$F(1,22) = 4.44, p = 0.05, \eta_p^2 = 0.17$] were significant (see **Table 4**).

To explain the triple interaction, further ANOVAs were performed including Type of Trial and Language as within-subject factors for the groups separately. In healthy individuals, no main effect nor interactions were statistically significant [$F_s < 1$]. In patients, only the interaction between Type of Trial and Language was significant [$F(1,10) = 4.87, p = 0.05, \eta_p^2 = 0.35$]. *Post hoc* analysis showed that patients performed similarly in repeat ($M = 1949.81, SD = 92.91$ ms) and switch trials ($M = 1938.27$ ms, $SD = 128.88$ ms; $p = 0.80$) in their dominant language, but significantly slower in switch ($M = 1998.90$ ms, $SD = 104.61$) than repeat ($M = 1882.09$ ms, $SD = 76.80$) trials when they performed the task in their non-dominant language ($p = 0.04$). This result suggests that patients suffered switch cost in their non-dominant language whereas controls did not.

Accuracy

In the analysis with healthy controls, we found a main effect of Type of trial to be significant [$F(1,12) = 7.19, p = 0.02, \eta_p^2 = 0.37$], suggesting that participants were less accurate in switch ($M = 96.86\%$, $SD = 0.74$) than repeat ($M = 98.82\%$, $SD = 0.29$) trials (see **Table 4**). No other main effects or interactions were statistically significant.

In the analysis with both groups, the main effect of Type of trial was significant [$F(1,22) = 5.11, p = 0.03, \eta_p^2 = 0.21$], suggesting that participants were less accurate in switch ($M = 93.91\%$, $SD = 1.32$) than repeat ($M = 96.43\%$, $SD = 1.53$) trials. Also, the main effect of Group was significant [$F(1,22) = 4.09, p = 0.05, \eta_p^2 = 0.17$] indicating that patients

($M = 92.79\%$, $SD = 1.71$) were less accurate than healthy controls ($M = 97.84\%$, $SD = 1.52$). No other main effect or interaction was statistically significant (see **Table 4**).

Flanker Task

Repeated-measures ANOVAs were performed including Type of Trial (congruent vs. incongruent) as a within-subject factor and Group (patients with aphasia vs. healthy controls) as a between-subject factor for RTs and accuracy separately. RTs were analyzed for correct responses only. Moreover, RTs across all conditions exceeding three standard deviations above or below mean were excluded from the analyses for each participant.

Reaction Times

The main effect of Type of Trial was significant [$F(1, 22) = 1191.73, p < 0.001, \eta_p^2 = 0.85$], suggesting than participants were slower in incongruent ($M = 990.33$ ms, $SD = 285.6$) than in congruent ($M = 879.64$ ms, $SD = 278.52$) trials. Also, the main effect of group was significant [$F(1,22) = 28.31, p < 0.001, \eta_p^2 = 0.57$], indicating that patients with aphasia were slower ($M = 1148.32$ ms, $SD = 285.74$) than healthy controls ($M = 711.09$ ms, $SD = 86.19$) to perform the task. Finally, the interaction between Type of Trial and Group was not statistically significant [$F(1,22) = 2.11, p = 0.17$], suggesting that the magnitude of the conflict cost was the same for the groups (see **Table 5**).

Accuracy

The main effect of Type of trial was significant [$F(1,22) = 7.05, p = 0.01, \eta_p^2 = 0.25$], suggesting higher accuracy in congruent ($M = 98.97\%$, $SD = 4.14$) than in incongruent trials ($M = 97.56\%$, $SD = 2.31$). However, no significant difference was found between patients with aphasia ($M = 97.12\%$, $SD = 3.07$) and healthy controls ($M = 99.26\%$, $SD = 1.45$) [$F(1, 22) = 2.78, p = 0.11$].

Correlations Between Linguistic and Non-linguistic Measures

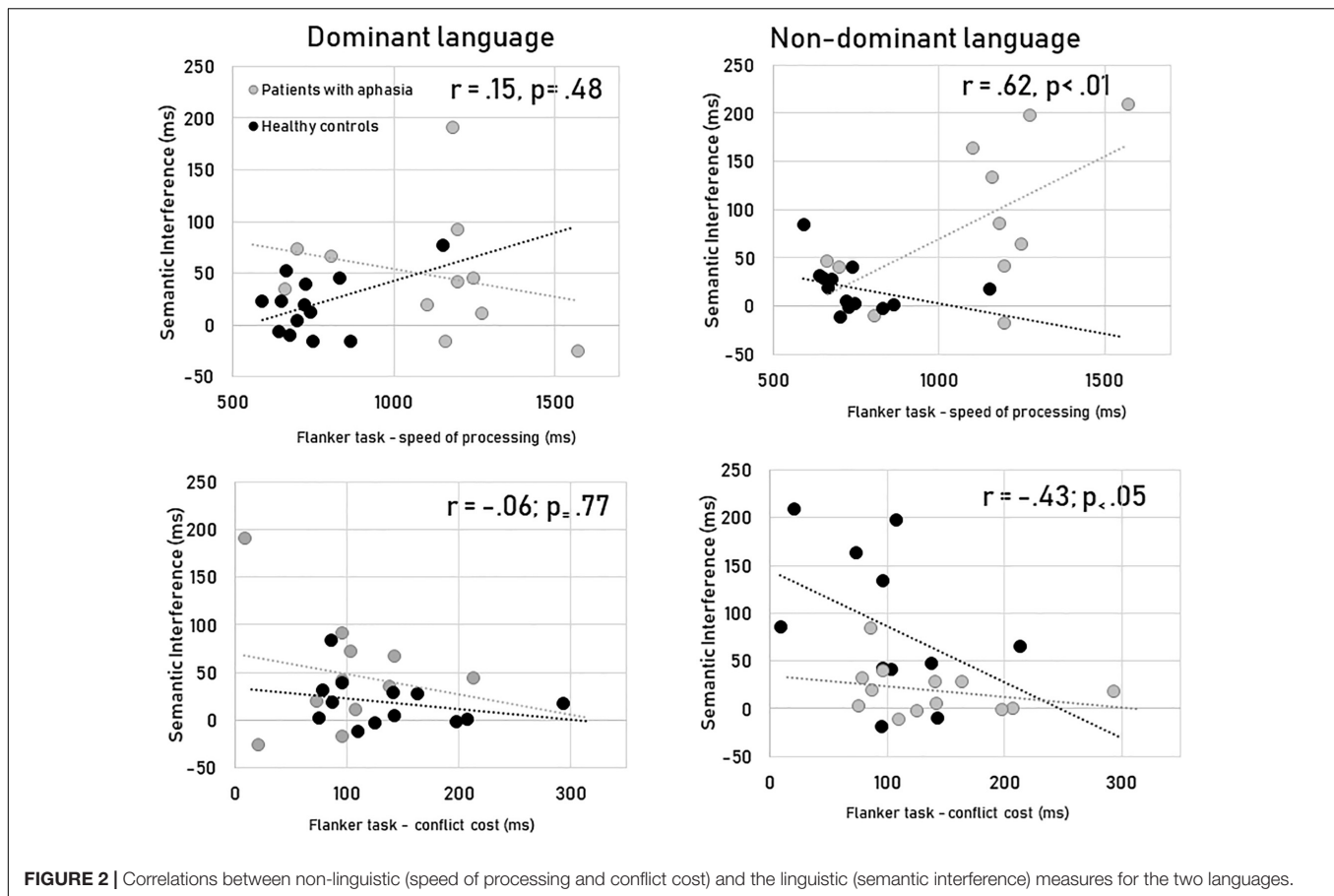
To address one of our hypotheses that language deficits might be related to non-linguistic control deficits, we performed correlations between each individual's performance on the tasks used to assess both domains.

TABLE 4 | RTs and accuracy in the bilingual word-picture matching task for healthy controls and patients with aphasia.

	Healthy controls				Patients with aphasia			
	Repeat	SE	Switch	SE	Repeat	SE	Switch	SE
RTs								
Dominant language	1051	61	1059	82	1949	66	1882	90
Non-dominant language	1041	64	1052	87	1938	70	1998	95
Accuracy								
Dominant language	98.8	1.5	96.5	2.7	94.1	1.6	89.8	2.9
Non-dominant language	98.6	1.1	97.0	2.1	94.1	1.2	92.6	2.2

TABLE 5 | RTs and accuracy in the flanker task for healthy controls and patients with aphasia.

Type of trial	Healthy controls		Patients with aphasia	
	Means	SE	Means	SE
RTs				
Congruent	649	47	1087	51
Incongruent	772	45	1200	49
Conflict costs	123		113	
Accuracy				
Congruent	99.9	0.7	98.0	0.6
Incongruent	98.6	1.2	96.5	1.2



For the non-linguistic domain, we used the individual speed of processing (congruent and incongruent trials) and the magnitude of the conflict cost (RTs on incongruent trials minus RTs on congruent trials) on the flanker task. For the linguistic domain, we used the magnitude of the semantic interference effect (RTs in homogeneous blocks minus RTs in heterogeneous block) and switch costs at the individual level for both dominant and non-dominant languages within the semantic blocked cyclic naming task and the bilingual matching task, respectively.

For the dominant language, the magnitude of the semantic interference did not correlate with the speed of processing [$r(24) = 0.15, p = 0.48$] and the conflict cost of the flanker task [$r(24) = -0.06; p = 0.77$] (see **Figure 2**). The switch costs in their dominant language did correlate with the speed of processing [$r(24) = 0.89, p < 0.001$], but not with the cost seen in the flanker task [$r(24) = -0.13; p = 0.55$].

However, for the non-dominant language, the magnitude of the semantic interference did correlate with the speed of processing [$r(24) = 0.62, p = 0.001$] and the conflict cost of the flanker task [$r(24) = -0.43; p = 0.05$]. The switch costs for the non-dominant language correlated with the speed of processing [$r(24) = 0.86, p < 0.001$] but not with the cost seen in the flanker task [$r(24) = -0.26; p = 0.22$].

Moreover, the degree of language impairment as indexed by the AQ of the WAB did not correlate with either non-linguistic

[speed of processing: $r(10) = -0.35, p = 0.16$; conflict cost: $r(10) = 0.19, p = 0.61$] or linguistic performance in patients for both languages on the semantic blocked naming task [dominant language: $r(10) = -0.45, p = 0.14$; non-dominant language: $r(10) = 0.41, p = 0.24$] and on the bilingual matching task [dominant language: $r(10) = 0.09, p = 0.82$; non-dominant language: $r(10) = 0.15, p = 0.70$].

DISCUSSION

With this study we aimed to investigate the language dependency of semantic processing in bilinguals. To address this question, we explored the performance of bilinguals with fluent aphasia and parallel language impairment on tasks of production and comprehension within their two languages. Furthermore, we used an EC task to explore whether the control mechanisms in the linguistic and non-linguistic domain may overlap.

We found three main results. First, semantic control processes related to lexical selection are language-dependent, as measured by a larger semantic interference effect during non-dominant language production in bilingual patients with aphasia. Second, the retrieval of semantic representations might have also a certain degree of language dependency under specific conditions, such as dual language contexts. Third, the linguistic processes of semantic

control show only partial overlap with those of domain-general EC (i.e., during conflict monitoring).

Language Dependency of Semantic Control in Production

We found evidence of language dependency for the semantic control system in bilinguals. Our bilingual patients with aphasia showed a higher semantic interference effect than healthy controls and, interestingly, to a greater degree when they did the task in their non-dominant language. First, it is important to stress that this result cannot be attributed to an imbalance of proficiency in the two languages in patients. In studies such as ours, it becomes necessary to exclude this variable as one of the factors that could explain differences in semantic processing between languages in bilingual patients with aphasia (Kiran and Edmonds, 2004; Lorenzen and Murray, 2008; Kiran and Roberts, 2010; Kiran et al., 2013a,b; Khachatryan et al., 2016). Given that patients in our study acquired their non-dominant language early on and had a similar frequency of usage in both their languages before injury, this possible confounding factor of language proficiency cannot account for the greater semantic interference in the non-dominant language observed.

Our results regarding semantic control complement previous studies that investigated the network of semantic representation in bilinguals. As reviewed in the Introduction, most of the neuroimaging studies have shown that bilinguals use very similar neural networks while they process semantic features of their L1 and L2 (e. g., Illes et al., 1999; Chee et al., 2001; Pillai et al., 2003; Grogan et al., 2009; Correia et al., 2014; Van de Putte et al., 2017). Similarly, neuropsychological studies of bilingual patients with semantic memory impairment indicate a comparable decline of the two languages, suggesting a common and shared neural network in the temporal lobe (Mendez et al., 1999; Hernández et al., 2008, 2010). However, it is important to highlight that, in the case of bilingual aphasia, we do not expect a deficit in semantic memory at the representational level, but rather a deficit in the control components of semantic retrieval (Jefferies and Lambon Ralph, 2006; Noonan et al., 2010). Similar to some extent to the previous concept of “access,” semantic control is in charge of retrieving the semantic information needed for a specific context and depends upon cognitive demand (Jefferies et al., 2008). Given this distinction, semantic control would be more affected in patients having lesions in fronto-temporoparietal areas due to decreased capacity to inhibit semantic competitors while their semantic representations could be spared (Jefferies and Lambon Ralph, 2006; Jefferies et al., 2008; Rogers et al., 2015). Therefore, we believe that our patients relied more on these control processes, within the linguistic domain, while they named pictures in those semantically blocked conditions where they were required to inhibit competitors. This type of competitive process has consequences at the lexical level, during selection and retrieval of the words. Accordingly, although we manipulated the degree of semantic competition within our task, we cannot exclude a possible effect at the lexical level since it is interconnected with the semantic units.

Different hypotheses have been proposed for the pathological effects found in patients when they have to name elements within semantically homogeneous conditions (Schnur et al., 2006, 2009; Harvey and Schnur, 2015). Our results seem to suggest that the problem experienced by aphasic patients in reducing semantic competition possibly comes from an excessive inhibition of lexical representations (McCarthy and Kartsounis, 2000). The semantic similarity between items would cause an increased level of inhibition on non-target words that would spread throughout the network; this same inhibition would then make a following, semantically related lexical item less accessible. Indeed, patients with aphasia showed more omissions than semantic errors, supporting the notion that they were not able to retrieve the correct name because it was completely inhibited. Moreover, this inhibitory process seems to be within-language since our patients did not produce many cross-language intrusions. This interpretation is more compatible with our findings than other hypotheses that proposed over-activation at the semantic level that builds up across cycles (Belke et al., 2005; Schnur et al., 2006; for a non-competitive selection account see Navarrete et al., 2014).

Interestingly, patients' ability to inhibit competitors during lexical retrieval was especially reduced while they were performing the task in their non-dominant language. This is not to say that the semantic representations of their non-dominant language were more affected. Rather, in control-demanding situations such as naming in their non-dominant language during homogenous conditions, lexical retrieval engaged the control network of semantic cognition to a greater degree and, in turn, resulted in a slowing down of the process. These results could be explained by some of the models of bilingual language production that have proposed language-dependency of lexico-semantic processing. For instance, Kroll and Stewart (1994) asserted that the lexico-semantic connections between L2 and L1 are weaker than those between L1 and L2; Gollan et al. (2008) also claimed that difference in frequency of language usage might explain why L2 retrieval is more demanding for bilinguals. However, Kroll and Stewart (1994)'s proposal is mainly based on data with late bilinguals and the predictions of their model are not entirely applicable to the population of early bilinguals that we studied (for a critical discussion of this issue, see Hernández et al., 2010). The only way to interpret our results with the Kroll and Stewart (1994)'s model would be to assume difference in the level of activation for lexical competition of the two languages (Kroll et al., 2010).

Language Dependency of Semantic Control in Comprehension

Differential language impairment observed in speech production also extended to comprehension. Our main aim was originally to study semantic processing during word production, but we decided to also include a comprehension task to check the integrity of the semantic representations. Although the matching task was not designed to measure the blocking and cycling effects of semantic interference, we found that when the two languages are mixed in a semantic matching task, the non-dominant is more affected than the dominant one. Conversely, healthy

controls did not show any switch cost and this is probably due to the nature of task. The studies that have used semantic categorization and lexical decision in healthy individuals have showed that switch costs are not always reliable in the matching tasks or they are reduced compared to production, probably because they require less involvement of language activation and inhibition (e.g., Orfanidou and Sumner, 2005; Macizo et al., 2012; Mosca and de Bot, 2017). We might say that this result in patients at the comprehension level could be related to some deficits in the access of lexical representations for the non-dominant language. Following brain damage, the competitive process (possibly of inhibitory nature) for lexical selection in the non-dominant language could be affected and this would explain why patients are more impaired in that language. Previous neuroimaging studies in bilinguals found mixed results: some that the control network described for language production (Abutalebi and Green, 2007, 2016; Calabria et al., 2018) is also active during word comprehension and recognition tasks (Peeters et al., 2019), but some other studies suggest that the overlap between the two system is only partial (Abutalebi et al., 2007; Blanco-Elorrieta and Pyllkänen, 2017).

These results show that the language-dependent nature of semantic control processes in bilinguals with aphasia during word production in single language contexts extends to word comprehension in dual language contexts. However, caution is required when interpreting these results due to important methodological differences between the two tasks. Despite the fact that the bilingual word-picture matching task could also be defined as a semantic task, participants performed it in a dual language condition, whereas, in the semantic blocked naming task, the two languages were not mixed. Future research should examine whether semantic control processes continue to exhibit a language-dependent nature during word comprehension when tested in single language contexts.

Semantic Control and EC in Bilinguals

The ‘executive selection account’ proposes that the effect of interference generated in the semantic blocked cyclic naming task is mediated by the involvement of domain-general EC mechanisms that are outside of the linguistic domain (Wilshire and McCarthy, 2002). In fact, there is evidence that performance on the Stroop task correlates with naming latencies in homogeneous conditions, suggesting that inhibition at response selection level would be the same in EC and semantic control (Crowther and Martin, 2014). Similarly, the involvement of the left inferior frontal gyrus and the left caudate nucleus has been interpreted as the EC network being responsible for resolving interference of semantic competitors (Canini et al., 2016). Given this evidence, we included the flanker task to measure individual domain-general EC performance.

Our results partially support this account. We found a positive correlation between the magnitude of the semantic interference effect and the speed of processing in the flanker task, but only for the non-dominant language. There was also a negative correlation between semantic interference and conflict cost, suggesting that a reduced magnitude of semantic interference is associated with smaller conflict costs. This observation is likely

biased by patients’ performance: given they are already very slow to respond in the congruent conditions, their “reduction” in conflict cost might be reflecting this generalized slowness rather than a true decrease in cost. In any case, this result seems to indicate that semantic competition does not overlap with general-domain EC mechanisms responsible for conflict resolution, contrary to what other studies have suggested (Crowther and Martin, 2014; Canini et al., 2016). Moreover, our patients were not impaired in conflict resolution as they had similar levels of conflict costs as healthy controls. However, they were generally slower, suggesting that the EC deficit they likely possessed was in conflict monitoring (Botvinick et al., 1999; Botvinick et al., 2001; Yeung, 2013). Conflict monitoring allows for the detection of potentially conflicting situations and subsequent adjustment of behavior when there is a switch from non-conflict situations (congruent trials) to conflict ones (incongruent trials) and vice versa. Therefore, the positive correlation between semantic interference and speed of processing may be interpreted in terms of an overlap of monitoring abilities (or deficits in these abilities for aphasic patients) between the linguistic domain (semantic control) and non-linguistic EC. Conflict resolution has been related to frontal activity (anterior cingulate cortex, Botvinick et al., 2001; for evidence in bilinguals see Abutalebi et al., 2012) and it is possible that this region was spared in our patients since they have fluent aphasia, a type of language disorder more strongly related to temporo-parietal damage. Therefore, we have to acknowledge the possibility that they did not show a deficit in conflict resolution because they did not have brain damage in frontal areas. Indeed, we know that higher semantic interference effects in bilingual patients are related to EC deficits when they have brain damage extending to the inferior frontal gyrus, as shown in a study by Schnur et al. (2009).

Previous studies that have compared linguistic to non-linguistic performance using a flanker task in bilingual patients with aphasia (Green et al., 2010; Verreyt et al., 2013; Dash and Kar, 2014; Gray and Kiran, 2015) have shown that there is an incomplete overlap between the two control systems. Further studies should explore other EC components, such as working memory and switching abilities, to determine whether these non-linguistic control mechanisms are more closely related to language control deficits in bilingual speakers.

CONCLUSION

The results of our study suggest that semantic control may be language-dependent and selective language impairment could be explained by an excessive inhibition placed upon the lexical representations of the non-dominant language. Additionally, semantic interference seems to be at least somewhat related to conflict monitoring deficits, suggesting a certain degree of overlap between EC and semantic control.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the “Parc de Salut MAR - Research Ethics

Committee” with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the “Parc de Salut MAR - Research Ethics Committee” (reference number: 2018/8029/I).

AUTHOR CONTRIBUTIONS

MC conceived the research, designed the experiments, and analyzed the data. NG, CS, and MS collected the experimental and neuropsychological data. Moreover, all the authors equally contributed to writing of the manuscript and discussion of the results.

FUNDING

MC was supported by the postdoctoral Ramón y Cajal fellowship (RYC-2013-14013) and Agencia Estatal de Investigación (AEI, National Research Agency) and Fondo Europeo de Desarrollo

Regional (FEDER, European Regional Development Fund) under project PSI2017-87784-R. This work was also supported by grants from the Catalan government (2017 SGR 268 and 2009 SGR 1521) and the European Union’s Seventh Framework Program for Research (No. 613465).

ACKNOWLEDGMENTS

We would like to dedicate this publication in memory of AC. His scientific contributions have been of great importance to the field of bilingualism and we are infinitely grateful to him for having been part of this study as one of the authors.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnhum.2019.00205/full#supplementary-material>

REFERENCES

- Abutalebi, J., Brambati, S. M., Annoni, J.-M., Moro, A., Cappa, S. F., and Perani, D. (2007). The neural cost of the auditory perception of language switches: an event-related functional magnetic resonance imaging study in bilinguals. *J. Neurosci.* 27, 13762–13769. doi: 10.1523/JNEUROSCI.3294-07.2007
- Abutalebi, J., Della Rosa, P. A., Green, D. W., Hernandez, M., Scifo, P., Keim, R., et al. (2012). Bilingualism tunes the anterior cingulate cortex for conflict monitoring. *Cereb. Cortex* 22, 2076–2086. doi: 10.1093/cercor/bhr287
- Abutalebi, J., and Green, D. W. (2007). Bilingual language production: the neurocognition of language representation and control. *J. Neurolinguist.* 20, 242–275.
- Abutalebi, J., and Green, D. W. (2016). Neuroimaging of language control in bilinguals: neural adaptation and reserve. *Biling. Lang. Cogn.* 19, 1–10. doi: 10.1017/S1366728916000225
- Belke, E. (2017). The role of task-specific response strategies in blocked-cyclic naming. *Front. Psychol.* 7:1955. doi: 10.3389/fpsyg.2016.01955
- Belke, E., Meyer, A. S., and Damian, M. F. (2005). Refractory effects in picture naming as assessed in a semantic blocking paradigm. *Q. J. Exp. Psychol.* A 58, 667–692. doi: 10.1080/02724980443000142
- Biegler, K., Crowther, J. E., and Martin, R. (2008). Consequences of an inhibition deficit for word production and comprehension: evidence from the semantic blocking paradigm. *Cogn. Neuropsychol.* 25, 493–527. doi: 10.1080/02643290701862316
- Blanco-Elorrieta, E., and Pyllkänen, L. (2017). Bilingual language switching in the lab vs. in the wild: the spatio-temporal dynamics of adaptive language control. *J. Neurosci.* 37, 9022–9036.
- Botvinick, M., Nystrom, L. E., Fissell, K., Carter, C. S., and Cohen, J. D. (1999). Conflict monitoring versus selection-for-action in anterior cingulate cortex. *Nature* 402, 179–181. doi: 10.1038/46035
- Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., and Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychol. Rev.* 108, 624–652.
- Branzi, F. M., Calabria, M., and Costa, A. (2018). “Cross-linguistic/bilingual language production,” in *Oxford Handbook of Psycholinguistics*, 2nd Edn, eds S. A. Rueschemeyer and M. G. Gaskell (Oxford: Oxford University Press), 463–483.
- Calabria, M., Costa, A., Green, D. W., and Abutalebi, J. (2018). Neural basis of bilingual language control. *Ann. N. Y. Acad. Sci.* 1426, 221–235. doi: 10.1111/nyas.13879
- Canini, M., Della Rosa, P. A., Catricalà, E., Strijkers, K., Branzi, F. M., Costa, A., et al. (2016). Semantic interference and its control: a functional neuroimaging and connectivity study. *Hum. Brain Map.* 37, 4179–4196. doi: 10.1002/hbm.23304
- Caramazza, A., and Costa, A. (2000). The semantic interference effect in the picture-word interference paradigm: does the response set matter? *Cognition* 75, B51–B64.
- Chee, M. W. L., Hon, N., Lee, H. L., and Soon, C. S. (2001). Relative language proficiency modulates BOLD signal change when bilinguals perform semantic judgments. *NeuroImage* 13, 1155–1163. doi: 10.1006/nimg.2001.0781
- Correia, J., Formisano, E., Valente, G., Hausfeld, L., Jansma, B., and Bonte, M. (2014). Brain-based translation: fMRI decoding of spoken words in bilinguals reveals language-independent semantic representations in anterior temporal lobe. *J. Neurosci.* 34, 332–338. doi: 10.1523/JNEUROSCI.1302-13.2014
- Costa, A., Miozzo, M., and Caramazza, A. (1999). Lexical selection in bilinguals: do words in the bilingual’s two lexicons compete for selection? *J. Mem. Lang.* 41, 365–397.
- Crowther, J. E., and Martin, R. C. (2014). Lexical selection in the semantically blocked cyclic naming task: the role of cognitive control and learning. *Front. Hum. Neurosci.* 8:9. doi: 10.3389/fnhum.2014.00009
- Damian, M. F., and Als, L. C. (2005). Long-lasting semantic context effects in the spoken production of object names. *J. Exp. Psychol. Learn. Mem. Cogn.* 31, 1372–1384. doi: 10.1037/0278-7393.31.6.1372
- Damian, M. F., and Bowers, J. S. (2003). Locus of semantic interference in picture-word interference tasks. *Psychon. Bull. Rev.* 10, 111–117. doi: 10.3758/BF03196474
- Dash, T., and Kar, B. R. (2014). Bilingual language control and general purpose cognitive control among individuals with bilingual aphasia: evidence based on negative priming and flanker tasks. *Behav. Neurol.* 2014:679706. doi: 10.1155/2014/679706
- Dijkstra, T., and van Heuven, W. J. B. (2003). The architecture of the bilingual word recognition system: from identification to decision. *Biling. Lang. Cogn.* 5, 175–197. doi: 10.1017/S1366728902003012
- Faroqi-Shah, Y., Frymark, T., Mullen, R., and Wang, B. (2010). Effect of treatment for bilingual individuals with aphasia: a systematic review of the evidence. *J. Neurolinguist.* 23, 319–341. doi: 10.1016/j.jneuroling.2010.01.002
- Faroqi-Shah, Y., Sampson, M., Pranger, M., and Baughman, S. (2016). Cognitive control, word retrieval and bilingual aphasia: is there a relationship? *J. Neurolinguist.* 45, 95–109. doi: 10.1016/j.jneuroling.2016.07.001
- Finkbeiner, M., Gollan, T. H., and Caramazza, A. (2006). Lexical access in bilingual speakers: what’s the (hard) problem? *Biling. Lang. Cogn.* 9, 153–166. doi: 10.1017/S1366728906002501
- Forster, K. I., and Forster, J. C. (2003). DMDX: a windows display program with millisecond accuracy. *Behav. Res. Methods Instrum. Comput.* 35, 116–124.

- Francis, W. S. (1999). Cognitive integration of language and memory in bilinguals: semantic representation. *Psychol. Bull.* 125, 193–222. doi: 10.1037/0033-2909.125.2.193
- Francis, W. S. (2005). “Bilingual semantic and conceptual representation,” in *Handbook of Bilingualism: Psycholinguistic Approaches*, eds J. F. Kroll and A. M. B. de Groot (New York, NY: Oxford University Press), 251–267.
- Goodglass, H., and Kaplan, E. (1972). *The assessment of aphasia and related disorders*. Philadelphia: Lea and Febiger.
- Gollan, T. H., Montoya, R. I., Cera, C., and Sandoval, T. C. (2008). More use almost always a means a smaller frequency effect: aging, bilingualism, and the weaker links hypothesis. *J. Mem. Lang.* 58, 787–814. doi: 10.1016/j.jml.2007.07.001
- Gray, T., and Kiran, S. (2015). The relationship between language control and cognitive control in bilingual aphasia. *Biling. Lang. Cogn.* 19, 1–20. doi: 10.1017/S1366728915000061
- Green, D. W. (1986). Control, activation, and resource: a framework and a model for the control of speech in bilinguals. *Brain Lang.* 27, 210–223.
- Green, D. W., Grogan, A., Crinion, J., Ali, N., Sutton, C., and Price, C. J. (2010). Language control and parallel recovery of language in individuals with aphasia. *Aphasiology* 24, 188–209. doi: 10.1080/02687030902958316
- Grogan, A., Green, D. W., Ali, N., Crinion, J. T., and Price, C. J. (2009). Structural correlates of semantic and phonemic fluency ability in first and second languages. *Cereb. Cortex* 19, 2690–2698. doi: 10.1093/cercor/bhp023
- Harvey, D. Y., and Schnur, T. T. (2015). Distinct loci of lexical and semantic access deficits in aphasia: evidence from voxel-based lesion-symptom mapping and diffusion tensor imaging. *Cortex* 67, 37–58. doi: 10.1016/j.cortex.2015.03.004
- Hernández, M., Cano, A., Costa, A., Sebastian-Galles, N., Juncadella, M., and Gascon-Bayarri, J. (2008). Grammatical category-specific deficits in bilingual aphasia. *Brain Lang.* 107, 68–80. doi: 10.1016/j.bandl.2008.01.006
- Hernández, M., Costa, A., Caño, A., Juncadella, M., and Gascon-Bayarri, J. (2010). On the translation routes in early and highly proficient bilingual people: evidence from an individual with semantic impairment. *Aphasiology* 24, 141–169. doi: 10.1080/02687030902958266
- Illes, J., Francis, W. S., Desmond, J. E., Gabrieli, J. D. E., Glover, G. H., Poldrack, R., et al. (1999). Convergent cortical representation of semantic processing in bilinguals. *Brain Lang.* 70, 347–363. doi: 10.1006/brln.1999.2186
- Jefferies, E. (2013). The neural basis of semantic cognition: converging evidence from neuropsychology, neuroimaging and TMS. *Cortex* 49, 611–625. doi: 10.1016/j.cortex.2012.10.008
- Jefferies, E., and Lambon Ralph, M. A. (2006). Semantic impairment in stroke aphasia versus semantic dementia: a case-series comparison. *Brain* 129, 2132–2147. doi: 10.1093/brain/awl153
- Jefferies, E., Patterson, K., and Ralph, M. A. L. (2008). Deficits of knowledge versus executive control in semantic cognition: insights from cued naming. *Neuropsychologia* 46, 649–658. doi: 10.1016/j.neuropsychologia.2007.09.007
- Kertesz, A., and Pascual-Leone García, A. (1990). *Bateria de Afasias “Western” (The Western Aphasia Battery en Versión y Adaptación Castellana)*. Valencia: Nau Llibres Ed.
- Khachatryan, E., Vanhoof, G., Beyens, H., Goeleven, A., Thijs, V., and Van Hulle, M. M. (2016). Language processing in bilingual aphasia: a new insight into the problem. *Wiley Interdiscip. Rev. Cogn. Sci.* 7, 180–196. doi: 10.1002/wcs.1384
- Kiran, S., and Edmonds, L. A. (2004). Effect of semantic naming treatment on crosslinguistic generalization in bilingual aphasia. *Brain Lang.* 91, 75–77. doi: 10.1016/j.bandl.2004.06.041
- Kiran, S., Grasmann, U., Sandberg, C., and Miikkulainen, R. (2013a). A computational account of bilingual aphasia rehabilitation. *Bilingualism* 16, 325–342. doi: 10.1017/S1366728912000533
- Kiran, S., Sandberg, C., Gray, T., Ascenso, E., and Kester, E. (2013b). Rehabilitation in bilingual aphasia: evidence for within- and between-language generalization. *Am. J. Speech Lang. Pathol.* 22, S298–S309. doi: 10.1044/1058-0360(2013)12-0085
- Kiran, S., and Roberts, P. M. (2010). Semantic feature analysis treatment in spanish-english and french-english bilingual aphasia. *Aphasiology* 24, 231–261. doi: 10.1080/02687030902958365
- Klein, D., Milner, B., Zatorre, R. J., Meyer, E., and Evans, A. C. (2006). The neural substrates underlying word generation: a bilingual functional-imaging study. *Proc. Natl. Acad. Sci. U.S.A.* 92, 2899–2903. doi: 10.1073/pnas.92.7.2899
- Kong, A. P. H. (2016). *Analysis of Neurogenic Disordered Discourse Production*. New York, NY: Taylor & Francis Group.
- Kroll, J. F., Dussias, P. E., Bice, K., and Perrotti, L. (2015). Bilingualism, Mind, and Brain. *Ann. Rev. Linguist.* 1, 377–394. doi: 10.1146/annurev-linguist-030514-124937
- Kroll, J. F., and Stewart, E. (1994). Category interference in translation and picture naming: evidence for asymmetric connections between bilingual memory representations. *J. Mem. Lang.* 33, 149–174. doi: 10.1006/jmla.1994.1008
- Kroll, J. F., Van Hell, J. G., Tokowicz, N., and Green, D. W. (2010). The revised hierarchical model: a critical review and assessment. *Bilingualism* 13, 373–381. doi: 10.1017/S136672891000009X
- La Heij, W. (2005). “Selection processes in monolingual and bilingual lexical access,” in *Handbook of Bilingualism Psycholinguistic Approaches*, eds J. F. Kroll and A. M. B. de Groot (New York, NY: Oxford University Press), 289–307.
- Laine, M., and Martin, N. (2006). *Anomia Theoretical and Clinical Aspects*. London: Psychology Press.
- Lambon Ralph, M. A., Jefferies, E., Patterson, K., and Rogers, T. T. (2017). The neural and computational bases of semantic cognition. *Nat. Rev. Neurosci.* 18, 42–55. doi: 10.1038/nrn.2016.150
- Lorezen, B., and Murray, L. L. (2008). Bilingual aphasia: a theoretical and clinical review. *Am. J. Speech Lang. Pathol.* 17, 299–317. doi: 10.1044/1058-0360(2008)026
- Macizo, P., Bajo, T., and Cruz Martín, M. (2010). Inhibitory processes in bilingual language comprehension: evidence from Spanish-English interlexical homographs. *J. Mem. Lang.* 63, 232–244. doi: 10.1016/j.jml.2010.04.002
- Macizo, P., Bajo, T., and Paolieri, D. (2012). Language switching and language competition. *Second Lang. Res.* 28, 131–149. doi: 10.1177/0267658311434893
- Mahon, B. Z., Costa, A., Peterson, R., Vargas, K. A., and Caramazza, A. (2007). Lexical selection is not by competition: a reinterpretation of semantic interference and facilitation effects in the picture-word interference paradigm. *J. Exp. Psychol. Learn. Mem. Cogn.* 33, 503–535. doi: 10.1037/0278-7393.33.3.503
- McCarthy, R. A., and Kartsounis, L. D. (2000). Wobbly words: refractory anomia with preserved semantics. *Neurocase* 6, 487–497. doi: 10.1080/13554790008402719
- Mendez, M. F., Perryman, K. M., Pontón, M. O., and Cummings, J. L. (1999). Bilingualism and dementia. *J. Neuropsychiatry Clin. Neurosci.* 11, 411–412.
- Mirman, D., and Britt, A. E. (2004). What we talk about when we talk. *Philos. Trans. Royal Soc. B* 18, 675–697. doi: 10.1080/0950238042000260351
- Moreno, E. M., Rodríguez-Fornells, A., and Laine, M. (2008). Event-related potentials (ERPs) in the study of bilingual language processing. *J. Neurolinguist.* 21, 477–508. doi: 10.1016/j.jneuroling.2008.01.003
- Mosca, M., and de Bot, K. (2017). Bilingual language switching: production vs. recognition. *Front. Psychol.* 8:934. doi: 10.3389/fpsyg.2017.00934
- Navarrete, E., Del Prato, P., Peressotti, F., and Mahon, B. Z. (2014). Lexical selection is not by competition: evidence from the blocked naming paradigm. *J. Mem. Lang.* 76, 253–272. doi: 10.1016/j.jml.2014.05.003
- Navarrete, E., Prato, D., and Mahon, B. Z. (2012). Factors determining semantic facilitation and interference in the cyclic naming paradigm. *Front. Psychol.* 3:38. doi: 10.3389/fpsyg.2012.00038
- Noonan, K. A., Jefferies, E., Corbett, F., and Lambon Ralph, M. A. (2010). Elucidating the nature of deregulated semantic cognition in semantic aphasia: evidence for the roles of prefrontal and temporo-parietal cortices. *J. Cogn. Neurosci.* 22, 1597–1613. doi: 10.1162/jocn.2009.21289
- Noonan, K. A., Jefferies, E., Visser, M., and Lambon Ralph, M. A. (2013). Going beyond inferior prefrontal involvement in semantic control: evidence for the additional contribution of dorsal angular gyrus and posterior middle temporal cortex. *J. Cogn. Neurosci.* 25, 1824–1850. doi: 10.1162/jocn_a_00442
- Nozari, N., and Hepner, C. R. (2018). To select or to wait? The importance of criterion setting in debates of competitive lexical selection. *Cogn. Neuropsychol.* doi: 10.1080/02643294.2018.1476335 [Epub ahead of print].
- Orfanidou, E., and Sumner, P. (2005). Language switching and the effects of orthographic specificity and response repetition. *Mem. Cogn.* 33, 355–369. doi: 10.3758/BF03195323
- Paradis, M., and Libben, G. (1987). *The Assessment of Bilingual Aphasia*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Patterson, K., Nestor, P. J., and Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nat. Rev. Neurosci.* 8, 976–987. doi: 10.1038/nrn2277

- Peeters, D., Vanlangendonck, F., Rueschemeyer, S. A., and Dijkstra, T. (2019). Activation of the language control network in bilingual visual word recognition. *Cortex* 111, 63–73. doi: 10.1016/j.cortex.2018.10.012
- Perani, D., and Abutalebi, J. (2005). The neural basis of first and second language processing. *Curr. Opin. Neurobiol.* 15, 202–206. doi: 10.1016/j.conb.2005.03.007
- Perea, M., Duñabeitia, J. A., and Carreiras, M. (2008). Masked associative/semantic priming effects across languages with highly proficient bilinguals. *J. Mem. Lang.* 58, 916–930. doi: 10.1016/j.jml.2008.01.003
- Pillai, J. J., Araque, J. M., Allison, J. D., Sethuraman, S., Loring, D. W., Thiruvaiyaru, D., et al. (2003). Functional MRI study of semantic and phonological language processing in bilingual subjects: preliminary findings. *NeuroImage* 19, 565–576. doi: 10.1016/S1053-8119(03)00151-4
- Protopapas, A. (2007). CheckVocal: a program to facilitate checking the accuracy and response time of vocal responses from DMDX. *Behav. Res. Methods* 39, 859–862.
- Rogers, T. T., Patterson, K., Jefferies, E., and Lambon Ralph, M. A. (2015). Disorders of representation and control in semantic cognition: effects of familiarity, typicality, and specificity. *Neuropsychologia* 76, 220–239. doi: 10.1016/j.neuropsychologia.2015.04.015
- Runnqvist, E., Strijkers, K., Alario, F. X., and Costa, A. (2012). Cumulative semantic interference is blind to language: implications for models of bilingual speech production. *J. Mem. Lang.* 66, 850–869. doi: 10.1016/j.jml.2012.02.007
- Schnur, T. T., Schwartz, M. F., Brecher, A., and Hodgson, C. (2006). Semantic interference during blocked-cyclic naming: evidence from aphasia. *J. Mem. Lang.* 54, 199–227. doi: 10.1016/j.jml.2005.10.002
- Schnur, T. T., Schwartz, M. F., Kimberg, D. Y., Hirshorn, E., Coslett, H. B., and Thompson-Schill, S. L. (2009). Localizing interference during naming: convergent neuroimaging and neuropsychological evidence for the function of Broca's area. *Proc. Natl. Acad. Sci. U.S.A.* 106, 322–327. doi: 10.1073/pnas.0805874106
- Schwartz, M. F. (2014). Theoretical analysis of word production deficits in adult aphasia. *Philos. Trans. Royal Soc. B* 369:20120390. doi: 10.1098/rstb.2012.0390
- Scott, R. M., and Wilshire, C. E. (2010). Lexical competition for production in a case of nonfluent aphasia: converging evidence from four different tasks. *Cogn. Neuropsychol.* 27, 505–538. doi: 10.1080/02643294.2011.598853
- Siyambalapatiya, S., Chenery, H. J., and Copland, D. A. (2013). Lexical-semantic representation in bilingual aphasia: findings from semantic priming and cognate repetition priming. *Aphasiology* 27, 1302–1321. doi: 10.1080/02687038.2013.817521
- Snodgrass, J. G., and Vanderwart, M. (1980). A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *J. Exp. Psychol. Hum. Learn. Mem.* 6, 174–215.
- Travis, C. E., Torres Cacoullos, R., and Kidd, E. (2017). Cross-language priming: a view from bilingual speech. *Biling. Lang. Cogn.* 20, 283–298. doi: 10.1017/S1366728915000127
- Van de Putte, E., De Baene, W., Brass, M., and Duyck, W. (2017). Neural overlap of L1 and L2 semantic representations in speech: a decoding approach. *NeuroImage* 162, 106–116. doi: 10.1016/j.neuroimage.2017.08.082
- Verreyt, N., De Letter, M., Hemelsoet, D., Santens, P., and Duyck, W. (2013). Cognate effects and executive control in a patient with differential bilingual aphasia. *Appl. Neuropsychol.* 20, 221–230. doi: 10.1080/09084282.2012.753074
- Wilshire, C. E., and McCarthy, R. A. (2002). Evidence for a context-sensitive word retrieval disorder in a case of nonfluent aphasia. *Cogn. Neuropsychol.* 19, 165–186. doi: 10.1080/02643290143000169
- Yeung, N. (2013). "Conflict monitoring and cognitive control," in *The Oxford Handbook of Cognitive Neuroscience: The Cutting Edges*, eds K. N. Ochsner and S. Kosslyn (New York, NY: Oxford University Press).
- Zeelenberg, R., and Pecher, D. (2003). Evidence for long-term cross-language repetition priming in conceptual implicit memory tasks. *J. Mem. Lang.* 49, 80–94. doi: 10.1016/S0749-596X(03)00020-2

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Calabria, Grunden, Serra, García-Sánchez and Costa. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Concrete vs. Abstract Semantics: From Mental Representations to Functional Brain Mapping

Nadezhda Mkrtychian^{1*}, Evgeny Blagovechtchenski¹, Diana Kurmakaeva¹, Daria Gnedykh¹, Svetlana Kostromina¹ and Yury Shtyrov^{1,2}

¹ Laboratory of Behavioral Neurodynamics, St. Petersburg State University, Saint Petersburg, Russia, ² Department of Clinical Medicine, Center of Functionally Integrative Neuroscience (CFIN), Aarhus University, Aarhus, Denmark

The nature of abstract and concrete semantics and differences between them have remained a debated issue in psycholinguistic and cognitive studies for decades. Most of the available behavioral and neuroimaging studies reveal distinctions between these two types of semantics, typically associated with a so-called “concreteness effect.” Many attempts have been made to explain these differences using various approaches, from purely theoretical linguistic and cognitive frameworks to neuroimaging experiments. In this brief overview, we will try to provide a snapshot of these diverse views and relationships between them and highlight the crucial issues preventing this problem from being solved. We will argue that one potentially beneficial way forward is to identify the neural mechanisms underpinning acquisition of the different types of semantics (e.g., by using neurostimulation techniques to establish causal relationships), which may help explain the distinctions found between the processing of concrete and abstract semantics.

OPEN ACCESS

Edited by:

Ferdinand Binkofski,
RWTH Aachen Universität, Germany

Reviewed by:

Anna M. Borghi,
Sapienza University of Rome, Italy
Felix R. Dreyer,
Freie Universität Berlin, Germany

*Correspondence:

Nadezhda Mkrtychian
st048999@student.spbu.ru;
Nadezhda.a.mkrtychian@gmail.com

Received: 27 February 2019

Accepted: 17 July 2019

Published: 02 August 2019

Citation:

Mkrtychian N,
Blagovechtchenski E, Kurmakaeva D,
Gnedykh D, Kostromina S and
Shtyrov Y (2019) Concrete vs.
Abstract Semantics: From Mental
Representations to Functional Brain
Mapping.
Front. Hum. Neurosci. 13:267.
doi: 10.3389/fnhum.2019.00267

Keywords: concrete and abstract semantics, concreteness effect, mental representation, brain, memory trace, psycholinguistics, functional brain mapping

DEFINING CONCRETENESS AND ABSTRACTNESS

One can often encounter in the literature such terms as “concrete and abstract concepts,” “concrete and abstract words,” or “concrete and abstract semantics.” What is the difference? In psycholinguistic and cognitive frameworks, concepts may be termed as the knowledge about a particular category (Barsalou et al., 2003), as a combination of atomic units of information and meaningful relationships between those units (Payne et al., 2007), or as “a mental representation of a class or individual which deals with *what* is being represented and *how* that information is typically used during the categorization” (Smith, 1989, p. 502). Such mental (internal or cognitive) representations (Paivio, 1990) are widely investigated in cognitive psychology, psycholinguistics, philosophy of mind and related fields (Carruthers and Cummins, 1990), but often without a clear connection to neural representations, which are more commonly addressed in brain research, neuroscience and neuroimaging. It is believed that the most important concepts (called lexical concepts) have an expression in the language in the form of individual words (= are labeled by words; Margolis and Laurence, 1999) and are thereby “our representation of word meaning” (Murphy, 2002, p. 392). In this regard, in most concept studies, linguistic stimuli are used and thus the terms “concept,” “word semantics,” and “word meaning” are often used

interchangeably. Traditionally, words/concepts are subdivided into concrete and abstract types, and this distinction is considered in many contemporary psycholinguistic and cognitive studies. As often claimed, concrete concepts/words have clear references to material objects (e.g., *dog*, *house*), whereas references of abstract ones are not physical entities, but more complex mental states (e.g., *thought*, *happiness*), conditions (*uncertainty*), situations (*encounter*), and relationships (*employment*) (Borghgi and Binkofski, 2014). However, even this seemingly simple distinction is not unequivocal. For instance, Myachykov and Fischer (2019) have argued that, in addition to this phenomenological dimension of abstractness, there are also sensorimotor and contextual aspects, and the same word/concept may be both concrete or abstract depending on different dimensions. Sensorimotor and contextual dimensions are, in turn, determined by individual life experience of lexicon acquisition and usage. Therefore, one way to extricate from this tangle could be studying processing of novel words, whose meanings are not yet represented in the participants' minds. Such an approach may solve the problem of conceptual confusion – the first obstacle to establishing of clear links between theoretical descriptions and the brain mechanisms which underlie representations of these different knowledge types in the brain.

THEORETICAL ACCOUNTS

Research of concrete and abstract concepts has a long history; a landmark event in its modern period was Paivio's seminal article "Abstractness, imagery, and meaningfulness in paired-associate learning" (Paivio, 1965). Numerous behavioral experiments using lexical decision, recognition, word naming, and other behavioral tasks demonstrated that concrete concepts, in comparison with abstract ones, are better remembered (Schwanenflugel et al., 1992), recognized (Fliessbach et al., 2006), faster read and comprehended (Schwanenflugel and Shoben, 1983), and faster learnt (Mestres-Missé et al., 2014). Similar results were revealed with respect to the processing of concrete and abstract verbs (Alyahya et al., 2018) and definitions (Borghgi and Zarccone, 2016). This advantage of concrete over abstract semantics is usually called "concreteness effect"; to help explain it, Paivio suggested the so-called dual-coding theory (DCT, Paivio, 1990) which posits two functional systems associated with semantic memory: verbal-based and imagery-based (non-verbal). These representational systems are interrelated and can be active independently or in parallel. According to DCT, whereas the verbal system may be responsible for coding both concrete and abstract concepts linguistically, the non-verbal imagery system is primarily involved in coding concrete – but not abstract – concepts, enhancing their processing and leading to behaviorally observed advantages (Kuiper and Paivio, 1977).

Notably, some investigations showed that concrete words elicit faster responses in lexical decision task only when there is no context information helping to understand the meaning; when context is available, the concreteness effect is reduced or absent (Schwanenflugel and Shoben, 1983). These observations

were explained by the context-availability theory (CAT), which claims that concrete and abstract concepts have different amount of semantic associations: concrete concepts have stronger associative connections with fewer contexts, while abstract concepts have weaker associative connections with a larger number of contexts. This, in turn, means that providing relevant context information may eliminate the "concreteness effect" leading to equally efficient processing of both semantic types.

A similar view on the distinctions between concrete and abstract words suggests that they are represented in mind in qualitatively different ways (Crutch, 2006). This hypothesis was based on the study of different types of semantic errors in patients with deep dyslexia. According to this account, concrete words have hierarchical semantic structure, which relies on categorical interrelationship (superordinate and co-ordinate), whereas abstract representations have, on the contrary, associative architecture (with connections between words commonly used together).

Other cognitive frameworks, rather than stressing the differences between abstract and concrete processing mechanisms, focus on searching for their similarities. For instance, the embodied cognition view on language grounds semantic representations in bodily functions (perception, action) and proposes that abstract word processing, in the same way as that of concrete words, relies, at least in part, on sensorimotor systems (Glenberg et al., 2008; Pulvermüller, 2013, see Borghgi et al., 2017, for review of embodied views on concrete/abstract concepts). Indeed, a comparison of acquisition and processing of abstract semantics in children with typical language development, atypical development, and autism showed no significant differences between these groups, also indicating the absence of specific mechanisms of abstract knowledge acquisition (Vigliocco et al., 2018). This, however, still does not exclude a more substantial contribution of the linguistic system into the abstract processing found in some studies (e.g., Sakreida et al., 2013).

In cognitive linguistics, a somewhat similar approach is offered by the so-called conceptual metaphor theory (CMT), an influential theoretical framework, according to which abstract concepts may be understood in reference to more concrete words by using metaphors (Lakoff and Johnson, 1980). However, in development, metaphors become available later than basic abstract knowledge as such; furthermore, it has been argued that not every abstract concept can be fully understood metaphorically, i.e., in terms of concrete words (Borghgi and Zarccone, 2016).

One theoretically contentious issue in accounting for concrete and abstract features of word semantics is that of a relationship between "concreteness" and "emotionality". Many authors consider words connected to emotions (for example, *love*, *joy*, *fear*) as a kind of abstract concepts (see, e.g., Dreyer and Pulvermüller, 2018) because they lack specific subject-relatedness. However, consideration of abstractness from embodied, rather than purely phenomenological dimension allows referring to emotions as concrete (embodied in individual experience) items (Myachykov and Fischer, 2019). Furthermore, some authors divide all concepts into three types: concrete,

abstract and emotional (Altarriba and Bauer, 2004). This latter approach seems somewhat controversial, as it does not appear to be based on uniform classification criteria. Moreover, both concrete and abstract words may possess less or more emotional meaning (consider, e.g., *joy* vs. *justice*, or *cake* vs. *pencil*); further, this may depend on a person's individual experience. To put it differently, it is uncertain why, in the Altarriba and Bauer (2004) classification, such words as *win* or *jeopardy* were included into the group of abstract words while *daughter* and *dentist* were treated as concrete words, even though their meaning clearly carries emotional aspects.

Perhaps a more convincing approach links emotional experience with abstract concepts (Kousta et al., 2011). The so-called affective grounding hypothesis (AGH) makes several specific suggestions in this respect (Lenci et al., 2018). First, abstract and concrete concepts differ in the extent of involvement of two types of information: experiential (sensory, motor, and affective) and linguistic (verbal associations); this clearly resonates with Paivio's dual-coding account. Second, concrete concepts are mainly grounded in sensory-motor information, whereas abstract word meanings are underpinned predominantly by linguistic and emotional information. Finally, the prevalence of these specific types of information plays a crucial role in acquisition as well as further representation of both concrete and abstract concepts (Vigliocco et al., 2009). As a side note, this approach provides a way to define specific semantics as a flexible combination of experiential and linguistic features, suggesting that abstractness and concreteness are relative terms, and not a simple binary distinction.

This view is complemented by a suggestion about a significant role of social experience in acquisition and representation of abstract concepts (Barsalou and Wiemer-Hastings, 2005), since linguistic experience is acquired directly or indirectly in social interactions which makes it particularly crucial in building up abstract knowledge. Borghi et al. (2018) support this idea, considering words as social tools (WAT theory) and suggesting that abstract representations are more likely to involve linguistic and social experience than concrete ones (because of the absence of material references with objects), especially during their acquisition (Borghi and Binkofski, 2014; Borghi and Zarcone, 2016). WAT is an attempt to create an integral theory of abstract concepts from the point of embodied and grounded approach to cognition. We concur with Borghi et al. (2018)' on the importance of exploring the differences between concrete and abstract concept acquisition but emphasize the need to focus on the dynamics of this acquisition process, not just on its outcomes.

NEUROSCIENTIFIC APPROACHES

A different avenue for disentangling various accounts and interpretations of cognitive phenomena is offered in neuroscience, which focuses on identifying their underlying brain mechanisms, by investigating neuroanatomical substrates and neurophysiological dynamics of cognitive processes in the brain. In simple terms, if comprehension of concrete and abstract

concepts is underpinned by different brain mechanisms, this can be investigated by scrutinizing neural activation patterns using functional brain mapping (e.g., EEG, MEG, fMRI or PET), or, to address causality, using neurostimulation techniques (TMS, tDCS) and/or brain-damaged patients. Neuropsychological data indicate that concrete words are more resistant to different brain injuries than abstract ones (Binder et al., 2005), suggesting at least partially different neural systems supporting these knowledge types. This suggestion is corroborated by a number of neuroscientific studies showing overlapping but not identical brain areas involved in abstract vs. concrete stimulus processing (see Montefinese, 2019, for a concise review).

However, there are still multiple contradictions across available neuroimaging studies (Wang et al., 2010), which has so far prevented neuroscience from resolving the dispute between theoretical accounts. Greater activation in such areas as middle and superior temporal gyrus (STG, MTG) and left inferior frontal gyrus (IFG) was associated with the processing of abstract concepts (Binder et al., 2005; Sabsevitz et al., 2005; Fliessbach et al., 2006; Pexman et al., 2007). Concrete concepts, in turn, have been shown to activate ventral anterior part of the fusiform gyrus (Sabsevitz et al., 2005; Bedny and Thompson-Schill, 2006; Fliessbach et al., 2006), which was also confirmed in an fMRI study of concrete word acquisition (Mestres-Misse et al., 2007). Other areas exhibit a less clear picture. For example, enhanced activation for abstract, as opposed to concrete, concepts has been observed in the anterior temporal region (ATL) in a number of studies (Tettamanti et al., 2008; Binder et al., 2009; Wang et al., 2010), whereas other experiments revealed the opposite, activation in ventral ATL specific for concrete concepts (Peelen and Caramazza, 2012; Visser et al., 2012; Robson et al., 2014), or an equal involvement of ventrolateral ATL for both concept types (Hoffman et al., 2015). It appears that while some such studies do not always have a clear basis in theoretical cognitive accounts, others mainly set out to prove the dual-coding theory. For instance, the results of EEG studies by Holcomb et al. (1999) speak in favor of the context-extended version of dual-coding account, which integrates DCT and CAT, at the neurophysiological level. Their experiments showed significant differences between brain responses to concrete and abstract words for the N400 component, a negative ERP wave associated with lexico-semantic processing: word concreteness leads to a greater negativity of the N400, especially in anterior areas, decreasing over posterior sites (Holcomb et al., 1999). Similar concreteness effect – stronger N400 – was also found in a study of acquisition of novel concrete and abstract semantics (Palmer et al., 2013). Concrete words also elicit larger N700 responses comparing with abstract ones even if they are matched for their context-availability and imageability (Barber et al., 2013), which, as the authors asserted, could not be explained by context-extended DCT. In turn, Pexman et al. (2007) unambiguously concluded that their neurophysiological data favors Barsalou's theory of semantic representation over dual-coding and context-availability theories, while Borghi et al. (2018) find neurophysiological support for the WAT theory, further deepening the theoretical divide.

There are virtually no studies of concrete vs. abstract semantics using brain stimulation techniques (which could provide the much-needed causal evidence), with only a handful of TMS papers that suggested prefrontal and motor areas to take part in abstract word comprehension (e.g., Vukovic et al., 2017). One way to apply brain stimulation is to investigate changes in the activity of the motor cortex and corticospinal activation during comprehension (Hoffman et al., 2010). For example, the processing of abstract and concrete phrases differentially modulates cortico-spinal excitability (Scorolli et al., 2012). However, any association with movement will cause activation of the mirror neurons in the motor system (Rizzolatti and Sinigaglia, 2016), and given the great variability in motor cortex responses to TMS (Fedele et al., 2016), it is very difficult to disentangle the specific and non-specific effects of different semantic types and brain stimulation.

Clinical data distinguishing between abstract and concrete concepts are extremely rare. While there are some cases (“case studies”) of specific impairments in abstract or concrete concept comprehension, separately, they are based on very limited observations ranging from one to four patients at most (Warrington and Crutch, 2005; Crutch, 2006; Tree and Kay, 2006). Furthermore, such conceptual comprehension impairments are confounded by a variety of other co-morbidities (e.g., dyslexia in Crutch, 2006), while the definitions of abstractness and concreteness used by the authors vary and do not always conform to the status quo in the field. In essence, the available clinical data are so far unable to provide a clear picture of distinctions between these semantic types.

FUTURE OUTLOOK

Whereas cognitive accounts of semantic representations, abstract semantics in particular, have gone a long way in recent decades, their neural counterparts so far suffer from the lack of studies and contradictions in the available data. The reasons for these contradictions could be many and include different properties of stimulus materials used, stimulation parameters, imaging modalities, and experimental tasks. One key difficulty lies with balancing basic psycholinguistic and physical properties of abstract and concrete words under investigations in a particular study; the lack of such balance confounds any differential results. A related issue that appears important is that most studies deal with pre-existing representations that are confounded by their surface properties, previous learning trajectories, daily use, and existing associations, all of which may obscure the results. In addition, the concrete-abstract dichotomy may not be complete and more fine-grain distinctions have been suggested: for example, action-related and object-/visually related concrete words, mental state-, emotion-, and mathematics-related abstract words (Dreyer and Pulvermüller, 2018). Further, rather than a dichotomy, there may be a multidimensional concreteness-abstractness continuum, along which words may vary, sometimes falling into both categories depending on the specific context (Myachykov and Fischer, 2019).

One way to circumvent these difficulties could be to assess the *process of acquisition* of novel concrete and abstract

semantics in laboratory settings, using stimuli with fully controlled and systematically modulated semantic, physical and psycholinguistic parameters. By observing the learning process behaviorally and its counterparts in the brain, it may be possible to elucidate the systems that take part in building up novel representations and the degree to which they differ between semantic types. To avoid confounds related to different modes of acquisition of abstract and concrete semantics, the learning regime should be maximally matched between semantic conditions, using, for instance, context-based inference or direct instruction (Atir-Sharon et al., 2015). To assess the learning outcomes, an elaborate testing of lexical, semantic, and contextual levels of acquisition is desirable; ideally, the assessment should be done both immediately and after a consolidation period (e.g., after an overnight sleep, Davis et al., 2009).

Whereas many acquisition studies use either exceptionally novel word forms (pseudowords) (De Groot and Keijzer, 2000; Mestres-Missé et al., 2014) or unfamiliar words of foreign languages with established semantics (van Hell and Mahn, 1997), it is crucial to disentangle the mechanisms of learning the new word form and its phonology from those of acquiring the semantics *per se* (Partanen et al., 2017). This, in our view, is best achieved by training well-matched phonologically and phonotactically legal forms both as such (i.e., surface forms only) *and* in conjunction with novel semantics – rather than attaching familiar semantics to novel native word forms or foreign words (Leminen et al., 2016).

There is still a predominance of studies dedicated to investigation of learning mechanisms of concrete rather than abstract semantics; they are targeted more often owing to their more obvious link with sensorimotor experience (Mahon and Caramazza, 2008) that lends itself readily to experimental manipulation. While it may be straightforward to learn new names for new objects using, e.g., word-picture matching, creating a new abstract category in an experimental setting is much more challenging. One way to address this could be adopting abstract concepts from cultures other than that of experimental participants.

On another note, most available studies use correlational measures, e.g., showing distinct activation patterns accompanying perception. Clearly, causal evidence is also needed to demonstrate functional relevance of such distinctions. Outside of limited patient studies, such evidence is presently lacking. The use of neurostimulation techniques (such as TMS or tDCS) to influence both comprehension and acquisition of concrete and abstract semantics may provide the much-needed evidence for the involvement of particular brain areas in representing specific semantic types. For example, Fiori et al. (2011) revealed that the application of anodal tDCS over Wernicke’s area while learning new words significantly improved the accuracy and decreased latencies in a picture-naming task, while another study (Flöel et al., 2008) showed faster and better associative verbal learning with the anodal tDCS over posterior left perisylvian areas, compared to sham. We are not aware of any similar studies comparing concrete and abstract semantics and their acquisition; this could be the target for future investigations.

To conclude, the literature suggests cognitively and neurophysiologically distinct systems that support abstract and concrete representations in mind and brain. Yet, the data available to date, particularly with respect to abstract semantics, do not allow for clear delineation of the underlying brain systems and thus explaining the effects found behaviorally. To fill these gaps in the field, future studies should use a combination of rigorously matched behavioral regimes, controlled modes of presentation, a comprehensive set of tasks to assess behavioral outcomes at different times, and different neuroimaging tools able to assess both the complex dynamics of word comprehension and the causal relationships between brain structures and representation types. One way to help disentangle the mechanisms underpinning different semantic

representations is to focus on their acquisition in controlled experimental settings.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

This study was supported by the RF Government grant contract no. 14.W03.31.0010.

REFERENCES

- Altarriba, J., and Bauer, L. M. (2004). The distinctiveness of emotion concepts: a comparison between emotion, abstract, and concrete words. *Am. J. Psychol.* 117, 389–410. doi: 10.2307/4149007
- Alyahya, R. S. W., Halai, A. D., Conroy, P., and Ralph, M. A. L. (2018). The behavioral patterns and neural correlates of concrete and abstract verb processing in aphasia: a novel verb semantic battery. *NeuroImage Clin.* 17, 811–825. doi: 10.1016/j.nicl.2017.12.009
- Atir-Sharon, T., Gilboa, A., Hazan, H., Koilis, E., and Manevitz, L. M. (2015). Decoding the formation of new semantics: MVPA investigation of rapid neocortical plasticity during associative encoding through fast mapping. *Neural Plast.* 2015:804385. doi: 10.1155/2015/804385
- Barber, H. A., Otten, L. J., Kousta, S.-T., and Vigliocco, G. (2013). Concreteness in word processing: ERP and behavioral effects in a lexical decision task. *Brain Lang.* 125, 47–53. doi: 10.1016/j.bandl.2013.01.005
- Barsalou, L. W., Simmons, W. K., Barbey, A. K., and Wilson, C. D. (2003). Grounding conceptual knowledge in modality-specific systems. *Trends Cogn. Sci.* 7, 84–91. doi: 10.1016/s1364-6613(02)00029-3
- Barsalou, L. W., and Wiemer-Hastings, K. (2005). “Situating abstract concepts,” in *Grounding Cognition: The Role of Perception and Action in Memory, Language, and Thinking*, eds D. Pecher and R. A. Zwaan (Cambridge: Cambridge University Press), 129–163. doi: 10.1017/CBO9780511499968.007
- Bedny, M., and Thompson-Schill, S. L. (2006). Neuroanatomically separable effects of imageability and grammatical class during single-word comprehension. *Brain Lang.* 98, 127–139. doi: 10.1016/j.bandl.2006.04.008
- Binder, J. R., Desai, R. H., Graves, W. W., and Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cereb. Cortex* 19, 2767–2796. doi: 10.1093/cercor/bhp055
- Binder, J. R., Westbury, C. F., McKiernan, K. A., Possing, E. T., and Medler, D. A. (2005). Distinct brain systems for processing concrete and abstract concepts. *J. Cogn. Neurosci.* 17, 905–917. doi: 10.1162/0899829054021102
- Borghi, A. M., Barca, L., Binkofski, F., Castelfranchi, C., Pezzulo, G., and Tummolini, L. (2018). Words as social tools: language, sociality and inner grounding in abstract concepts. *Phys. Life Rev.* doi: 10.1016/j.plrev.2018.12.001 [Epub ahead of print].
- Borghi, A. M., and Binkofski, F. (2014). *Words as Social Tools: An Embodied View on Abstract Concepts*. New York, NY: Springer, doi: 10.1007/978-1-4614-9539-0
- Borghi, A. M., Binkofski, F., Castelfranchi, C., Cimatti, F., Scorolli, C., and Tummolini, L. (2017). The challenge of abstract concepts. *Psychol. Bull.* 143, 263–292. doi: 10.1037/bul0000089
- Borghi, A. M., and Zarcone, E. (2016). Grounding abstractness: abstract concepts and the activation of the mouth. *Front. Psychol.* 7:1498. doi: 10.3389/fpsyg.2016.01498
- Carruthers, P., and Cummins, R. (1990). Meaning and mental representation. *Philos. Q.* 40, 527–530. doi: 10.2307/2220119
- Crutch, S. J. (2006). Qualitatively different semantic representations for abstract and concrete words: further evidence from the semantic reading errors of deep dyslexic patients. *Neurocase* 12, 91–97. doi: 10.1080/13554790500507172
- Davis, M. H., Di Betta, A. M., Macdonald, M. J. E., and Gaskell, M. G. (2009). Learning and consolidation of novel spoken words. *J. Cogn. Neurosci.* 21, 803–820. doi: 10.1162/jocn.2009.21059
- De Groot, A. M. B., and Keijzer, R. (2000). What is hard to learn is easy to forget: the roles of word concreteness, cognate status, and word frequency in foreign-language vocabulary learning and forgetting. *Lang. Learn.* 50, 1–56. doi: 10.1111/0023-8333.00110
- Dreyer, F. R., and Pulvermüller, F. (2018). Abstract semantics in the motor system? – An event-related fMRI study on passive reading of semantic word categories carrying abstract emotional and mental meaning. *Cortex* 100, 52–70. doi: 10.1016/j.cortex.2017.10.021
- Fedele, T., Blagovechtchenski, E., Nazarova, M., Iscan, Z., Moiseeva, V., and Nikulin, V. V. (2016). Long-range temporal correlations in the amplitude of alpha oscillations predict and reflect strength of intracortical facilitation: combined TMS and EEG study. *Neuroscience* 331, 109–119. doi: 10.1016/j.neuroscience.2016.06.015
- Fiori, V., Coccia, M., Marinelli, C. V., Vecchi, V., Bonifazi, S., Ceravolo, M. G., et al. (2011). Transcranial direct current stimulation improves word retrieval in healthy and nonfluent aphasic subjects. *J. Cogn. Neurosci.* 23, 2309–2323. doi: 10.1162/jocn.2010.21579
- Fliessbach, K., Weis, S., Klaver, P., Elger, C. E., and Weber, B. (2006). The effect of word concreteness on recognition memory. *Neuroimage* 32, 1413–1421. doi: 10.1016/j.neuroimage.2006.06.007
- Flöel, A., Rösser, N., Michka, O., Knecht, S., and Breitenstein, C. (2008). Noninvasive brain stimulation improves language learning. *J. Cogn. Neurosci.* 20, 1415–1422. doi: 10.1162/jocn.2008.20098
- Glenberg, A. M., Sato, M., Cattaneo, L., Riggio, L., Palumbo, D., and Buccino, G. (2008). Processing abstract language modulates motor system activity. *Q. J. Exp. Psychol.* 61, 905–919. doi: 10.1080/17470210701625550
- Hoffman, P., Binney, R. J., and Lambon Ralph, M. A. (2015). Differing contributions of inferior prefrontal and anterior temporal cortex to concrete and abstract conceptual knowledge. *Cortex* 63, 250–266. doi: 10.1016/j.cortex.2014.09.001
- Hoffman, P., Jefferies, E., and Lambon Ralph, M. A. (2010). Ventrolateral prefrontal cortex plays an executive regulation role in comprehension of abstract words: convergent neuropsychological and repetitive TMS evidence. *J. Neurosci.* 30, 15450–15456. doi: 10.1523/JNEUROSCI.3783-10.2010
- Holcomb, P. J., Kounios, J., Anderson, J. E., and West, W. C. (1999). Dual-coding, context-availability, and concreteness effects in sentence comprehension: an electrophysiological investigation. *J. Exp. Psychol. Learn. Mem. Cogn.* 25, 721–742. doi: 10.1037/0278-7393.25.3.721
- Kousta, S.-T., Vigliocco, G., Vinson, D. P., Andrews, M., and Del Campo, E. (2011). The representation of abstract words: why emotion matters. *J. Exp. Psychol. Gen.* 140, 14–34. doi: 10.1037/a0021446

- Kuiper, N. A., and Paivio, A. (1977). Incidental recognition memory for concrete and abstract sentences equated for comprehensibility. *Bull. Psychonom. Soc.* 9, 247–249. doi: 10.3758/bf03336989
- Lakoff, G., and Johnson, M. (1980). *Metaphors We Live By*. Chicago, IL: University of Chicago Press.
- Leminen, A., Kimppe, L., Leminen, M. M., Lehtonen, M., Mäkelä, J. P., and Shtyrov, Y. (2016). Acquisition and consolidation of novel morphology in human neocortex: a neuromagnetic study. *Cortex* 83, 1–16. doi: 10.1016/j.cortex.2016.06.020
- Lenci, A., Lehani, G. E., and Passaro, L. C. (2018). The emotions of abstract words: a distributional semantic analysis. *Top. Cogn. Sci.* 10, 550–572. doi: 10.1111/tops.12335
- Mahon, B. Z., and Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *J. Physiol. Paris* 102, 59–70. doi: 10.1016/j.jphysparis.2008.03.004
- Margolis, E., and Laurence, S. (1999). *Concepts: Core Readings*. Cambridge, MA: MIT Press.
- Mestres-Missé, A., Münte, T. F., and Rodriguez-Fornells, A. (2014). Mapping concrete and abstract meanings to new words using verbal contexts. *Second Lang. Res.* 30, 191–223.
- Mestres-Misse, A., Rodriguez-Fornells, A., and Munte, T. F. (2007). Watching the brain during meaning acquisition. *Cereb. Cortex* 17, 1858–1866. doi: 10.1093/cercor/bhl094
- Montefinese, M. (2019). Semantic representation of abstract and concrete words: a minireview of neural evidence. *J. Neurophysiol.* 121, 1585–1587. doi: 10.1152/jn.00065.2019
- Murphy, G. L. (2002). *The Big Book of Concepts*. Cambridge, MA: MIT Press.
- Myachykov, A., and Fischer, M. H. (2019). A hierarchical view of abstractness: grounded, embodied, and situated aspects. *Phys. Life Rev.* (in press). doi: 10.1016/j.plrev.2019.04.005
- Paivio, A. (1965). Abstractness, imagery, and meaningfulness in paired-associate learning. *J. Verbal Learn. Verbal Behav.* 4, 32–38. doi: 10.1016/s0022-5371(65)80064-0
- Paivio, A. (1990). “Dual coding theory,” in *Mental Representations: A Dual Coding Approach*, eds D. E. Broadbent, J. L. McGaugh, N. J. Mackintosh, M. I. Posner, E. Tulving, and L. Weiskrantz (Oxford: Oxford University Press), 53–83. doi: 10.1093/acprof:oso/9780195066661.003.0004
- Palmer, S. D., MacGregor, L. J., and Havelka, J. (2013). Concreteness effects in single-meaning, multi-meaning and newly acquired words. *Brain Res.* 1538, 135–150. doi: 10.1016/j.brainres.2013.09.015
- Partanen, E., Leminen, A., de Paoli, S., Bundgaard, A., Kingo, O. S., Krøjgaard, P., et al. (2017). Flexible, rapid and automatic neocortical word form acquisition mechanism in children as revealed by neuromagnetic brain response dynamics. *Neuroimage* 155, 450–459. doi: 10.1016/j.neuroimage.2017.03.066
- Payne, P. R. O., Mendonça, E. A., Johnson, S. B., and Starren, J. B. (2007). Conceptual knowledge acquisition in biomedicine: a methodological review. *J. Biomed. Inform.* 40, 582–602. doi: 10.1016/j.jbi.2007.03.005
- Peelen, M. V., and Caramazza, A. (2012). Conceptual object representations in human anterior temporal cortex. *J. Neurosci.* 32, 15728–15736. doi: 10.1523/jneurosci.1953-12.2012
- Pexman, P. M., Hargreaves, I. S., Edwards, J. D., Henry, L. C., and Goodyear, B. G. (2007). Neural correlates of concreteness in semantic categorization. *J. Cogn. Neurosci.* 19, 1407–1419. doi: 10.1162/jocn.2007.19.8.1407
- Pulvermüller, F. (2013). How neurons make meaning: brain mechanisms for embodied and abstract-symbolic semantics. *Trends Cogn. Sci.* 17, 458–470. doi: 10.1016/j.TICS.2013.06.004
- Rizzolatti, G., and Sinigaglia, C. (2016). The mirror mechanism: a basic principle of brain function. *Nat. Rev. Neurosci.* 17, 757–765. doi: 10.1038/nrn.2016.135
- Robson, H., Zahn, R., Keidel, J. L., Binney, R. J., Sage, K., and Ralph, M. A. L. (2014). The anterior temporal lobes support residual comprehension in Wernicke’s aphasia. *Brain* 137, 931–943. doi: 10.1093/brain/awt373
- Sabsevitz, D. S., Medler, D. A., Seidenberg, M., and Binder, J. R. (2005). Modulation of the semantic system by word imageability. *Neuroimage* 27, 188–200. doi: 10.1016/j.neuroimage.2005.04.012
- Sakreida, K., Scorolli, C., Menz, M. M., Heim, S., Borghi, A. M., and Binkofski, F. (2013). Are abstract action words embodied? An fMRI investigation at the interface between language and motor cognition. *Front. Hum. Neurosci.* 7:125. doi: 10.3389/fnhum.2013.00125
- Schwanenflugel, P. J., Akin, C., and Luh, W.-M. (1992). Context availability and the recall of abstract and concrete words. *Mem. Cogn.* 20, 96–104. doi: 10.3758/bf03208259
- Schwanenflugel, P. J., and Shoben, E. J. (1983). Differential context effects in the comprehension of abstract and concrete verbal materials. *J. Exp. Psychol. Learn. Mem. Cogn.* 9, 82–102. doi: 10.1037/0278-7393.9.1.82
- Scorolli, C., Jacquet, P. O., Binkofski, F., Nicoletti, R., Tessari, A., and Borghi, A. M. (2012). Abstract and concrete phrases processing differentially modulates cortico-spinal excitability. *Brain Res.* 1488, 60–71. doi: 10.1016/j.brainres.2012.10.004
- Smith, E. E. (1989). “Concepts and induction,” in *Foundations of Cognitive Science*, ed. M. I. Posner (Cambridge, MA: The MIT Press), 501–526.
- Tettamanti, M., Manenti, R., Della Rosa, P. A., Falini, A., Perani, D., Cappa, S. F., et al. (2008). Negation in the brain: modulating action representations. *Neuroimage* 43, 358–367. doi: 10.1016/j.neuroimage.2008.08.004
- Tree, J. J., and Kay, J. (2006). Phonological dyslexia and phonological impairment: an exception to the rule? *Neuropsychologia* 44, 2861–2873. doi: 10.1016/j.neuropsychologia.2006.06.006
- van Hell, J. G., and Mahn, A. C. (1997). Keyword mnemonics versus rote rehearsal: learning concrete and abstract foreign words by experienced and inexperienced learners. *Lang. Learn.* 47, 507–546. doi: 10.1111/0023-8333.00018
- Vigliocco, G., Meteyard, L., Andrews, M., and Kousta, S. (2009). Toward a theory of semantic representation. *Lang. Cogn.* 1, 219–247. doi: 10.1515/langcog.2009.011
- Vigliocco, G., Ponari, M., and Norbury, C. (2018). Learning and processing abstract words and concepts: insights from typical and atypical development. *Top. Cogn. Sci.* 10, 533–549. doi: 10.1111/tops.12347
- Visser, M., Jefferies, E., Embleton, K. V., and Lambon Ralph, M. A. (2012). Both the middle temporal gyrus and the ventral anterior temporal area are crucial for multimodal semantic processing: distortion-corrected fMRI evidence for a double gradient of information convergence in the temporal lobes. *J. Cogn. Neurosci.* 24, 1766–1778. doi: 10.1162/jocn_a_00244
- Vukovic, N., Feurra, M., Shpektor, A., Myachykov, A., and Shtyrov, Y. (2017). Primary motor cortex functionally contributes to language comprehension: an online rTMS study. *Neuropsychologia* 96, 222–229. doi: 10.1016/j.neuropsychologia.2017.01.025
- Wang, J., Conder, J. A., Blitzer, D. N., and Shinkareva, S. V. (2010). Neural representation of abstract and concrete concepts: a meta-analysis of neuroimaging studies. *Hum. Brain Mapp.* 31, 1459–1468. doi: 10.1002/hbm.20950
- Warrington, E. K., and Crutch, E. J. (2005). Jargon dyslexia in an individual with semantic dementia: further evidence for task-specificity in phonological output. *Neurocase* 11, 351–362. doi: 10.1080/13554790591006375

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Mkrtychian, Blagovechtchenski, Kurmakaeva, Gnedyykh, Kostromina and Shtyrov. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Effects of Visual Priming and Event Orientation on Word Order Choice in Russian Sentence Production

Mikhail Pokhoday^{1*}, Yury Shtyrov^{1,2} and Andriy Myachykov^{1,3}

¹ Centre for Cognition and Decision Making, Institute of Cognitive Neuroscience, National Research University Higher School of Economics, Moscow, Russia, ² Department of Clinical Medicine, Center of Functionally Integrative Neuroscience, Aarhus University, Aarhus, Denmark, ³ Department of Psychology, Northumbria University, Newcastle upon Tyne, United Kingdom

OPEN ACCESS

Edited by:

Pia Knoeferle,
Humboldt University of Berlin,
Germany

Reviewed by:

L. Robert Slevc,
University of Maryland, College Park,
United States
Helene Kreysa,
Friedrich Schiller University Jena,
Germany

*Correspondence:

Mikhail Pokhoday
mpokhoday@hse.ru;
mikhail.pokhoday@gmail.com

Specialty section:

This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Psychology

Received: 20 February 2019

Accepted: 01 July 2019

Published: 20 August 2019

Citation:

Pokhoday M, Shtyrov Y and
Myachykov A (2019) Effects of Visual
Priming and Event Orientation on
Word Order Choice in Russian
Sentence Production.
Front. Psychol. 10:1661.
doi: 10.3389/fpsyg.2019.01661

Existing research shows that distribution of the speaker's attention among event's protagonists affects syntactic choice during sentence production. One of the debated issues concerns the extent of the attentional contribution to syntactic choice in languages that put stronger emphasis on word order arrangement rather than the choice of the overall syntactic frame. To address this, the current study used a sentence production task, in which Russian native speakers were asked to verbally describe visually perceived transitive events. Prior to describing the target event, a visual cue directed the participants' attention to the location of either the agent or the patient of the subsequently presented visual event. In addition, we also manipulated event orientation (agent-left vs. agent-right) as another potential contributor to syntactic choice. The number of patient-initial sentences was the dependent variable compared between conditions. First, the obtained results replicated the effect of visual cueing on the word order in Russian language: more patient-initial sentences in patient cued condition. Second, we registered a novel effect of event orientation: Russian native speakers produced more patient-initial sentences after seeing events developing from right to left as opposed to left-to-right events. Our study provides new evidence about the role of the speaker's attention and event orientation in syntactic choice in language with flexible word order.

Keywords: attention, constituent ordering, Russian language, perceptual priming, event orientation

INTRODUCTION

Every day we effortlessly produce sentences talking about objects, actions, people, and events. Producing sentences about visually perceived events requires several choices to be made by the speaker. Some of these choices refer to the selection of the syntactic structure of the produced sentence. When describing a transitive event for example, a speaker of English can choose between active and passive voice frames. In addition to the choice between structural alternatives, many languages offer their speakers the choice between different word-order options (scrambling; Gell-Mann and Ruhlen, 2011). These two processes relate to the question addressed in this paper: How does the speaker decide which particular frame to choose and how to arrange the constituents in a sentence? Here, we report the results of a sentence production study that investigated how

manipulations of visual attention and event orientation affect speakers' choice of word order in Russian – a free-order language that supports scrambling via explicit case marking and explicit constituent agreement.

In a visually situated context, the sentence production process begins with image apprehension. At this stage, input from perceptual modalities (e.g., visual, auditory, and motor) provides initial information for conceptual and linguistic interpretation of the event, with attention acting as a filter modulating and ranking the input according to what is relevant, noticeable, or important. The final product of this filtration process is then coded by the production system and is reflected in a generated sentence. Existing literature provides evidence that the speaker's attentional state is reflected in their choice of syntactic structure (see Myachykov et al., 2018b for a recent review). In one of the earliest studies (Tomlin, 1995), English-speaking participants watched a film depicting one fish (the agent) eating another fish (the patient). Attention of the speaker was manipulated by means of an explicit (i.e., consciously processed) exogenous visual cue – an arrow pointer above either the agent or the patient. The task was to continuously describe the interaction between the two fish including the eating event itself (the *target event*). Descriptions of the target events were analyzed for their syntactic structure: participants produced more active voice descriptions (e.g., *the blue fish eats the red fish*) when the cue was on the agent fish. When, however, attention was directed to the patient fish, a passive voice description (e.g., *the red fish was eaten by the blue fish*) was more likely. This and similar findings indicate that attention to one of the interacting protagonists is reflected in the sentence production strategies, which include assigning the referents to their constituent roles in the sentence (Gleitman et al., 2007; Myachykov et al., 2011, 2012a,b, 2018a; Coco and Keller, 2012, 2015; Iwabuchi et al., 2013; Montag and MacDonald, 2014; Rommers et al., 2017; Pokhoday and Myachykov, 2018; Pokhoday et al., 2018).

At the same time, it remains unclear whether the attentional contribution to structural choice is universal across languages. After all, English is a language with a largely restricted word order while other languages (Russian, Finnish, etc.) rely upon a wider degree of word-order flexibility. This question was addressed only in a couple of existing reports (Myachykov et al., 2011; Hwang and Kaiser, 2014). One study (Myachykov and Tomlin, 2008) used a methodology similar to Tomlin (1995) studying Russian native speakers. The results indicated that, unlike their English counterparts, Russian speakers did not assign the *subject* role to the cued referent; instead, they selected it as the sentential starting point generating patient-initial or agent-initial *active-voice* word orders in both cueing conditions. One explanation for this difference is a different degree of reliance on syntactic alternations and scrambling strategies in English and Russian: While syntactic alternations (e.g., active/passive) are quite common in English, Russian uses its explicit morphology, making scrambling a more productive and more frequently used mechanism (Kolomackiy, 2009).

While this finding provided initial evidence for the role of the speaker's attentional focus in Russian sentence production, it was confounded by methodological limitations similar to the

ones pointed out by Bock et al. (2004). The most critical points were (1) the repetitive use of the event of one fish eating the other in all trials without filler materials, (2) the explicitness of the cueing manipulation – the parallel presentation of the cue and the target. In real-life communication, salience, including visual salience, can be much more subtle; hence, one may need to use equally subtle attention manipulations in order to properly understand the role of attentional focus in structural choice. In English, such modifications have been implemented in studies that successfully replicated the original findings by Tomlin using improved experimental designs (e.g., Gleitman et al., 2007; Myachykov et al., 2012a, 2018a, as well as by authors of this paper in Pokhoday et al., 2018). However, the same has never been done in studies investigating the role of attention in sentence production in flexible word-order languages.

Another important contributor to the speaker's behavior that rarely features in sentence production studies is the asymmetry of event conceptualization. Naturally, the same event can be perceived from a variety of perspectives that have little to do with the event's salience but rather reflect speakers' top-down biases. Some of these top-down biases have been extensively studied. For example, conceptual accessibility – or “the ease with which the mental representation of some potential referent can be activated in or retrieved from memory” (Bock and Warren, 1985, p. 50) has been shown to bias structural choices in a manner very similar to that of attention – a more accessible referent tends to be assigned a more prominent grammatical role in a produced sentence. Individual components that were shown to increase conceptual accessibility and bias syntactic choice include referential imageability (Bock and Warren, 1985), givenness (Bock, 1977; Arnold et al., 2000), animacy (Prat-Sala and Branigan, 2000; Christianson and Ferreira, 2005; Altmann and Kemper, 2006; Branigan et al., 2008), definiteness (Grieve and Wales, 1973), and prototypicality (Kelly et al., 1986).

Yet another top-down feature that biases speakers' conceptualization of the described event has to do with the distribution of the thematic roles among the event protagonists. More specifically, some reports suggest that the event's agent is more likely to be conceptualized ahead of the event's patient and be assigned a more prominent syntactic role, e.g., that of a Subject (Kemmerer, 2012; Cohn and Paczynski, 2013). This so-called “agent advantage” was supported in a recent study by Hafri et al. (2018). In their work they tested how the role of the referent character affects performance of participants in the unrelated tasks (attending to visual features unrelated to the roles). They found that if the target referent switched from agent to patient between trials, the response time increased. These authors concluded that such pattern of results reflects the automaticity and rapidness of referent role extraction during event perception. Overall, “other” and “error” accounted for less than 2% of the total responses (for full data see **Supplementary Table S1**).

The mental representations of the events tend to reflect the conceptualization asymmetry described above (Santiago et al., 2010; Tversky, 2011). Santiago et al. (2010), for example, investigated the direction of mental representations of perceived events. They reported results of three experiments, which indicate that participants perceived both video events and static

events on a continuum from left to right. Tversky (2011) also discussed the existence of canonical (agent on the left) and non-canonical (agent on the right) event representations. These findings suggest a degree of canonicity in event perception with the establishment of a top-down effect that can be traced in sentence production strategies. In addition, a study by Dobel et al. (2007) tested whether the event orientation effect is a result of a hemispheric specialization or a cultural preference. They compared the drawings of German (left-to-right reading and writing) and Hebrew (right-to-left reading and writing) speakers. Participants heard a sentence in which the position of agent or recipient has been manipulated, then they were to draw the event. Hebrew speakers draw left-to-right events positioning the agent on the left about 30% less frequently than German speakers. Dobel et al. (2007) concluded that there exists a bias consistent with a reading direction and thus supported the cultural hypothesis (see also Maass and Russo, 2003). Similarly, a study by Esaulova et al. (2018) had German and Arabic speakers describe visually presented events with the agent positioned on the left or on the right. Arabic speakers preferred to start their descriptions with the agents on the right while their German counterparts demonstrated the opposite preference. Hence, positioning of the referents in visual scenes may be shaped by the characteristics of the particular writing system used in the speakers' language.

Here, we address both aforementioned features – an improved control of attention in comparison with previous work and control of agent-patient asymmetry in event conceptualization – at once. In general, we predict that the left-to-right processing bias, common in left-to-right readers, will lead to faster processing and a higher probability of using the referent on the left as the sentential starting point. In addition, if event orientation is a significant contributor to syntactic choice, one would predict an interaction between the cue location and event orientation (Myachykov et al., 2007). In sum, the present study aimed at testing the degree of the perceptual visual priming effect in syntactic alternations during Russian transitive sentence production. Deeper investigation of that aspect of sentence production can hint at the existence of different language production mechanisms, in this case grammatical role assignment mechanism, between English and Russian.

METHODS

This experiment was approved by the Local Ethics Committee of the National Research University Higher School of Economics, Moscow.

Participants

To determine the sample size we used previous research as reference. 24 participants (18 females, mean age = 21, $SD = 1.62$) recruited from the students and staff population at the HSE University took part in the study. To participate in the study, participants had to be native Russian speakers, have normal (or corrected to normal) vision, and have no language or attention-related impairments (e.g., dyslexia and ADHD). Participants

received course credits or monetary remuneration for their participation. All participants gave written informed consent before taking part.

Design

We have adopted the procedure from our previous work (Myachykov et al., 2012a,b; Pokhoday et al., 2018). Two independent variables were manipulated: Cue Location (toward the *agent* or toward the *patient*) and Event Orientation (*Agent on the left* or *Agent on the right*). This resulted in a 2×2 factorial design with Cue Location and Event Orientation as within-subjects/within-items factors. The dependent variable was the proportion of the sentences where Patient referent was the first element of the sentence (Patient-first sentences).

Materials

To keep experimental conditions similar to our previous studies (Pokhoday et al., 2018) we have used the same stimulus materials [adopted from Myachykov et al. (2012a,b)]. Target pictures depicted six transitive events rotated between sixteen referents (see **Appendix 1** for the list of events and referents). We have crossed over the characters and the events to create 48 transitive-event target stimuli (**Figure 1** for example). Each event, performed by different characters, was shown to a participant eight times. Participants received an equal number of Left-to-Right and Right-to-Left stimuli pictures. Materials were presented in a pseudo-random order such that a minimum of two filler pictures separated target pictures from each other. Filler materials ($N = 96$) were included to avoid potential structural priming bias (e.g., Bock, 1986). In filler trials, participants described ditransitive or intransitive events. In ditransitive filler trials, participants produced either double-object or prepositional-object structures. In intransitive filler trials, they produced single-referent SV sentences. Materials were arranged into four lists, which allowed all events to feature in all four experimental conditions in a fully counterbalanced fashion. Each participant saw only one list out of four.

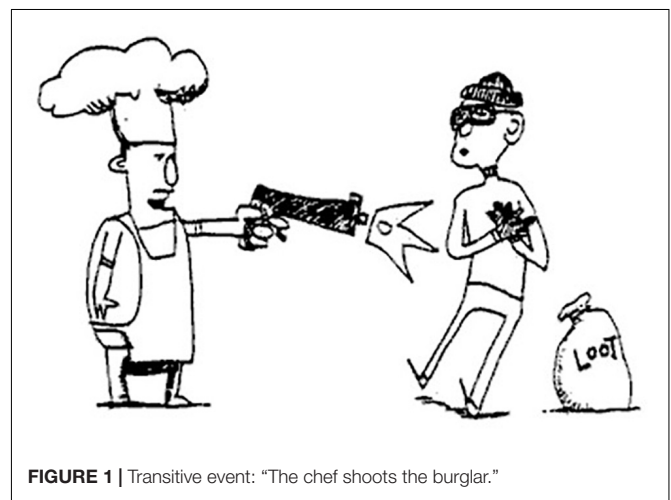


FIGURE 1 | Transitive event: "The chef shoots the burglar."

Apparatus

The experiment was created in SR Research Experiment Builder v2.1.140 software (SR Research Ltd., Ottawa, ON, Canada). An EyeLink 1000+ Desktop eye tracker (SR Research) was used to record fixation locations prior to presentation of a perceptual cue in order to avoid any possible directional biases. Eye movements were recorded from the right eye only with a 1000 Hz sampling rate. Stimuli were delivered by the eye tracker PC to an ASUS VG248QE 24-inch display (refresh rate 144 Hz). Generated sentences were recorded using a voice recorder application (Smart Recorder 1.8.0, SmartMob) and stored on a password protected PC. Participants were seated 60 cm away from the monitor with their head position controlled by a chinrest.

Procedure

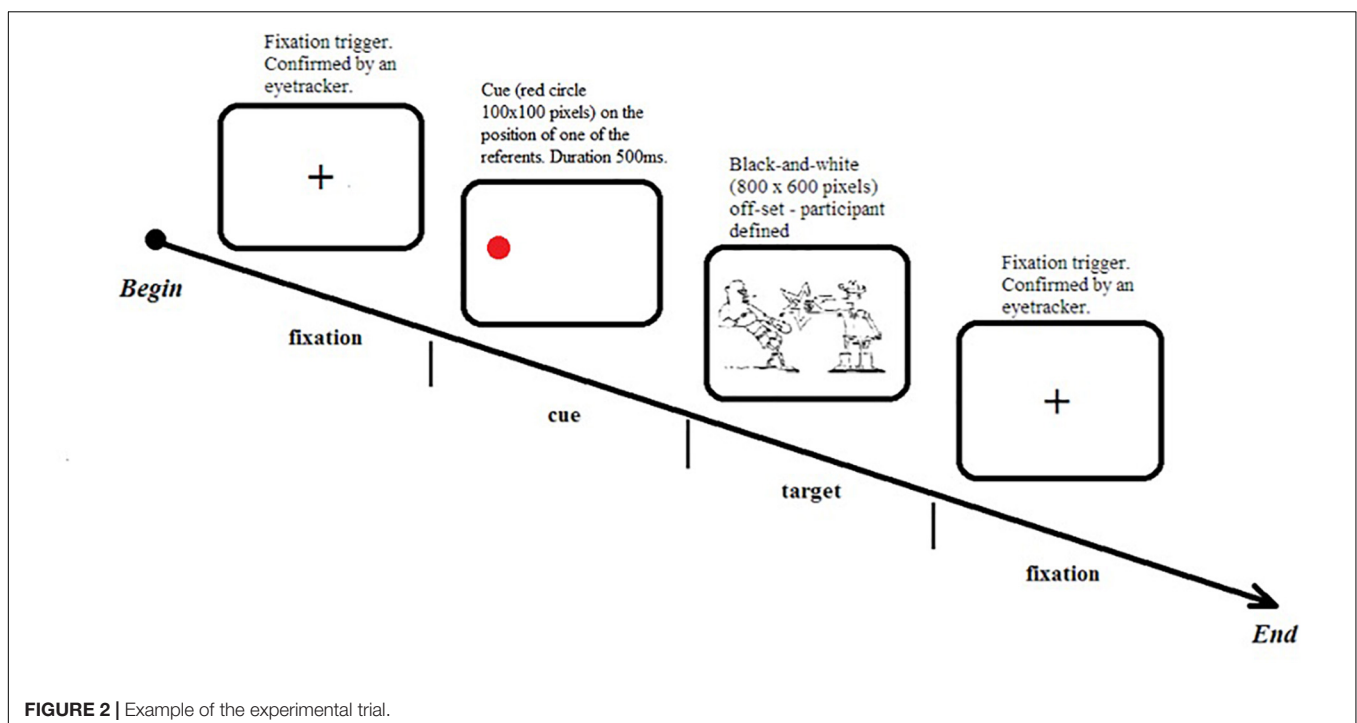
The study took place in the eye-tracking laboratory of the HSE Centre for Cognition and Decision Making. Before the experiment, participants provided their demographics and signed consent forms. After reading experimental instructions, participants received a practice session followed by the eye tracker calibration procedure (standard 9-point calibration, average calibration error 0.37°). The practice session consisted of two tasks. First, participants familiarized themselves with the 16 referents: the characters' depictions were sequentially presented centrally on screen, with their names written underneath. Participants' task was to read out loud and remember the character's names. This ensured that participants knew the referents' appearances and names in order to minimize cognitive effort related to recognizing the referents' identities and retrieving their names during the main experiment. This procedure also

helped to reduce potential ambiguity in naming referents [e.g., "маляр" (*painter*) – for the character "художник" (*artist*)]. Second, participants practiced describing events similar to the ones they would later encounter in the main experimental session. Participants saw fourteen randomly selected events in an individually randomized order, with each picture depicting an event with one or two referents (previously practiced) and the event's name in the infinitive form [e.g., "зваться" (*to chase*)] written underneath. As before, participants were instructed to examine the event and read its name aloud. The purpose of the event practice session was to minimize the variability of potential lexical candidates for the event description [e.g., "ударить" (*to strike*), for "бить" (*to hit*) event].

Upon completion of the practice session, participants received instructions for the main part of the experiment. Participants were told that every trial would begin with the presentation of a black cross in the middle of the screen (until fixation was confirmed by the eye tracker) followed by a red circle (the cue for 500 ms) in various locations, finally followed by the presentation of a picture stimulus (until participant pressed the space bar). The cue location corresponded to the subsequent position of one of the referents. Participants were instructed to look at the black cross, then, on appearance of the red circle, direct their gaze to it, wait for the event, and then describe the event aloud in one sentence mentioning both characters and their interaction. On completion of each trial (**Figure 2**), participants proceeded to the next trial by pressing the spacebar.

Data Analysis

The audio recordings of participants' responses were transcribed and responses were coded as follows: (0) Agent First or (1) Patient



First. Produced passive voice sentences ($N = 6$) were coded as Patient First sentences, as they were OVS. The responses that were not classifiable as (0) or (1) were coded as “other.” Erroneous and absent responses were coded as “error.” Overall, “other” and “error” accounted for less than 2% of the total responses.

According to the currently well-established practice we performed inferential analyses using Generalized Linear Mixed Effects Models (GLMM), as part of the lme4 package in R (R Core Team¹). The dependent variable of interest was the use of patient initial description (True = 1 and False = 0). A binary logistic model was specified in the family argument of the glmer() function. The model included a full-factorial Cue Location (Agent, Patient) \times Event Orientation (Left-to-right, Right-to-left) fixed effects design. All predictors were mean-centered using deviation-coding. We adopted the maximal random effects structure (Barr et al., 2013) justified by the design. We included in the model random correlations; by-subject and by-item random intercepts, by-subject and by-item random slopes for every main effect. These were included as both factors were within-subject and within-items. P -values were obtained via Likelihood Ratio Chi-Square ($LR\chi^2$) model comparisons.

RESULTS

Overall, 24 participants provided 1152 responses, 1131 of which were included into the analysis. The grand average intercept of the GLMM was estimated as -2.600 log odds units ($SE = 0.289$), which is well below zero (and in turn much smaller than 0.5 in probability space). Hence, patient-initial responses (13.5%) were greatly outnumbered by agent-initial responses (86.5%; see Table 1 for absolute counts), an expected result that is in line with previous experimental findings (Myachykov and Tomlin, 2008).

¹<https://www.R-project.org>

Figure 3 summarizes the distribution of the patient-initial responses across experimental conditions. It is clear that, overall, there were more patient-first sentences in the patient-cued conditions than in the agent-cued conditions. This was supported by a reliable main effect of Cue Location [$LR\chi^2(1) = 17.268$, $p < 0.001$]; the parameter estimations clarified that there were more patient-initial sentences when the patient referent was primed ($b = -0.845$, $SE = 0.200$, $p < 0.001$). We also registered the main effect of Event Orientation [$LR\chi^2(1) = 5.95$, $p = 0.01$]: there were more patient-initial responses when the agent was on the right side ($b = -0.500$, $SE = 0.198$, $p < 0.001$). Notably, there was no significant interaction between Cue Location and Event Orientation [$LR\chi^2(1) = 2.86$, $p = 0.09$; $b = -0.694$, $SE = 0.398$, $p = 0.08$].

In order to verify whether our sample size was adequate, we ran a *post hoc* observed power analysis. Results showed that this sample size was enough to register a moderate size priming effect (Mahowald et al., 2016). Considering the GLMM parameter estimates effect sizes of our factors of Cue location and Event Orientation were as log odds of -0.845 and -0.500 , respectively. Thus, the general odds ratio effect sizes for these effects were $\exp(0.845) = 2.32$ and $\exp(0.500) = 1.64$. Average syntactic priming effects with and without lexical overlap reported in

TABLE 1 | Probabilities of agent vs. patient responses across all participants and trials (absolute cell counts in brackets) by levels of event orientation (agent-left and agent-right) and cue location (agent and patient).

Event orientation	Cue location	Total	Agent initial	Patient initial
Agent-left	Agent	282	0.908 (256)	0.092 (26)
	Patient	279	0.868 (242)	0.132 (37)
Agent-right	Agent	285	0.899 (256)	0.101 (29)
	Patient	285	0.779 (222)	0.221 (63)

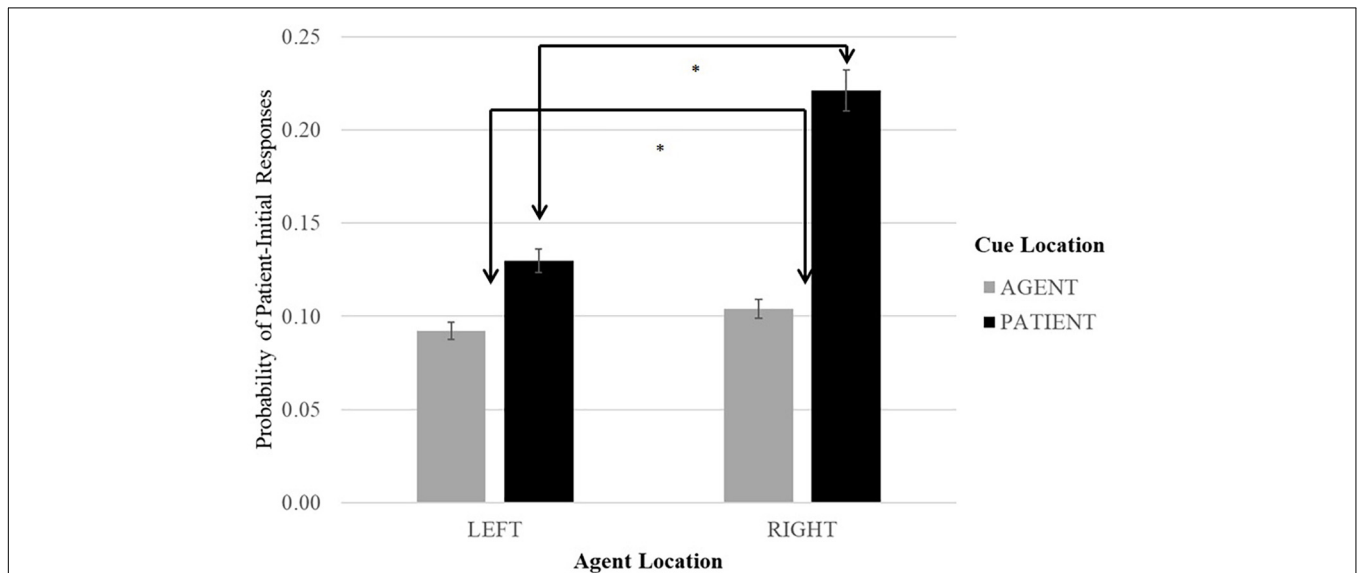


FIGURE 3 | Proportion of Patient-Initial responses. Error bars represent Standard Errors. *Significantly different.

Mahowald et al. (2016) are 3.26 and 1.67, respectively. So, our main effect sizes are within or very close to general benchmarks of similar studies.

DISCUSSION

In this study, we have investigated the combined effects of perceptual priming and event orientation on the speaker's word-order choices in Russian. Evidences suggest that perceptual priming of attention affects syntactic choice of the speaker. However, it is still unknown whether different word-order flexibility languages rely on similar mechanisms. Here, we collected data from Russian native speakers in order to assess the existence of perceptual priming effects on syntactic choice in Russian. Important addition in our study was the inclusion of event orientation in the analysis, which allowed us comparisons between bottom-up (cueing) and top-down (event orientation) priming effects. Below we discuss implications of our study.

First, we have replicated the previously reported perceptual priming effect (Myachykov and Tomlin, 2008) in a study with improved methodology and better experimental controls. We have also demonstrated that event orientation influenced syntactic choice via imposing an additional bias on the ordering of the constituents driven by the canonical left-to-right event scanning. The latter is evident as there were more patient-initial sentences when the agent was presented on the right side of the depicted event. According to some researchers, this effect might reflect the general left-to-right scanning mechanism associated with the automated writing and reading habits (e.g., Dobel et al., 2007; Santiago et al., 2010; Tversky, 2011; Esaulova et al., 2018). We did not register a reliable interaction between Cue Location and Event orientation, which suggests that the word-order choice in Russian can accommodate either the attentional (bottom-up) bias or the event orientation (top-down), but not both of these biases simultaneously. What can possibly happen is that the priming effect of the visual cue diminishes by the time structure coding occurs, while the priming effect of event orientation is present throughout all production stages due to the presence of the target stimuli picture throughout trial.

Overall, the results of the study support the hypothesis that perceptual priming influences constituent ordering but not the choice of syntactic structure in Russian. Passive-voice responses were almost non-existent in the patient-cued condition while participants still consistently encoded the cued referent as the initial element of the produced sentence. What is left unknown is whether this mechanism is similar to that of English language. As we have used

Patient-initial sentences in comparison to Passive voice sentences used in English language studies, the similarity of the implied mechanism is questionable and further research is therefore necessary. Another open question is which attention network is affecting syntactic choice? This may possibly be addressed by using an Attention Network Test (Fan et al., 2005) followed in combination with stimulation of the related brain areas.

DATA AVAILABILITY

All datasets generated for this study are included in the manuscript and/or the **Supplementary Files**.

ETHICS STATEMENT

The National Research University Higher School of the Economics Ethics Committee Participants have signed consent forms prior to experiment. At the end of the experiment they have been debriefed.

AUTHOR CONTRIBUTIONS

MP contributed to write up, data collection, data analysis, and hypothesis of the manuscript. YS contributed to reviewing, editing, hypothesis, and supervision of the manuscript. AM contributed to hypothesis, reviewing, editing, analysis, and supervision of the manuscript.

FUNDING

This manuscript was prepared within the framework of the HSE University Basic Research Program and funded by the Russian Academic Excellence Project "5-100."

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2019.01661/full#supplementary-material>

TABLE S1 | Russian language data of 24 participants. DV: amount of patient-initial sentences. IV: event orientation and visual cue location.

REFERENCES

- Altmann, L. J., and Kemper, S. (2006). Effects of age, animacy and activation order on sentence production. *Lang. Cogn. Process.* 21, 322–354. doi: 10.3389/fpsyg.2017.02246
- Arnold, J. E., Losongco, A., Wasow, T., and Ginstrom, R. (2000). Heaviness vs. newness: the effects of structural complexity and discourse status on constituent ordering. *Language* 76, 28–55. doi: 10.1353/lan.2000.0045
- Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: keep it maximal. *J. Mem. Lang.* 68, 255–278. doi: 10.1016/j.jml.2012.11.001
- Bock, J. K. (1977). The effect of a pragmatic presupposition on syntactic structure in question answering. *J. Verbal Learn. Verbal Behav.* 16, 723–734. doi: 10.1016/s0022-5371(77)80031-5
- Bock, J. K. (1986). Syntactic persistence in language production. *Cogn. Psychol.* 18, 355–387. doi: 10.1016/0010-0285(86)90004-6

- Bock, J. K., and Warren, R. K. (1985). Conceptual accessibility and syntactic structure in sentence formulation. *Cognition* 21, 47–67. doi: 10.1016/0010-0277(85)90023-x
- Bock, K., Irwin, D. E., and Davidson, D. J. (2004). “Putting first things first,” in *The Interface of Language, Vision, and Action: Eye Movements and the Visual World* (pp. 249–278), eds J. M. Henderson and F. Ferreira (New York, NY: Psychology Press).
- Branigan, H. P., Pickering, M. J., and Tanaka, M. (2008). Contributions of animacy to grammatical function assignment and word order during production. *Lingua* 118, 172–189. doi: 10.1016/j.lingua.2007.02.003
- Christianson, K., and Ferreira, F. (2005). Conceptual accessibility and sentence production in a free word order language (Odawa). *Cognition* 98, 105–135. doi: 10.1016/j.cognition.2004.10.006
- Coco, M. I., and Keller, F. (2012). Scan patterns predict sentence production in the cross-modal processing of visual scenes. *Cogn. Sci.* 36, 1204–1223. doi: 10.1111/j.1551-6709.2012.01246.x
- Coco, M. I., and Keller, F. (2015). The interaction of visual and linguistic saliency during syntactic ambiguity resolution. *Q. J. Exp. Psychol.* 68, 46–74. doi: 10.1080/17470218.2014.936475
- Cohn, N., and Paczynski, M. (2013). Prediction, events, and the advantage of agents: the processing of semantic roles in visual narrative. *Cogn. Psychol.* 67, 73–97. doi: 10.1016/j.cogpsych.2013.07.002
- Dobel, C., Diesendruck, G., and Bölte, J. (2007). How writing system and age influence spatial representations of actions: a developmental, cross-linguistic study. *Psychol. Sci.* 18, 487–491. doi: 10.1111/j.1467-9280.2007.01926.x
- Esaulova, Y., Penke, M., and Verlage, S. (2018). “Preferences for the positioning of actants in visual scenes,” in *Poster Presented at the Second International Conference, Cologne*.
- Fan, J., McCandliss, B. D., Fossella, J., Flombaum, J. I., and Posner, M. I. (2005). The activation of attentional networks. *Neuroimage* 26, 471–479. doi: 10.1016/j.neuroimage.2005.02.004
- Gell-Mann, M., and Ruhlen, M. (2011). The origin and evolution of word order. *Proc. Natl. Acad. Sci. U.S.A.* 108, 17290–17295. doi: 10.1073/pnas.1113716108
- Gleitman, L. R., January, D., Nappa, R., and Trueswell, J. C. (2007). On give and take between event apprehension and utterance formulation. *J. Mem. Lang.* 57, 544–569. doi: 10.1016/j.jml.2007.01.007
- Grieve, R., and Wales, R. J. (1973). Passives and topicalization. *Br. J. Psychol.* 64, 173–182. doi: 10.1111/j.2044-8295.1973.tb01340.x
- Hafri, A., Trueswell, J. C., and Strickland, B. (2018). Encoding of event roles from visual scenes is rapid, spontaneous, and interacts with higher-level visual processing. *Cognition* 175, 36–52. doi: 10.1016/j.cognition.2018.02.011
- Hwang, H., and Kaiser, E. (2014). Having a syntactic choice is not always better: the effects of syntactic flexibility on Korean production. *Lang. Cogn. Neurosci.* 29, 1115–1131. doi: 10.1080/23273798.2013.875212
- Iwabuchi, T., Inui, T., Ohba, M., and Ogawa, K. (2013). A functional MRI study of a picture–sentence verification task: evidence of attention shift to the grammatical subject. *NeuroReport* 24, 298–302. doi: 10.1097/WNR.0b013e32835f8826
- Kelly, M. H., Bock, J. K., and Keil, F. C. (1986). Prototypicality in a linguistic context: effects on sentence structure. *J. Mem. Lang.* 25, 59–74. doi: 10.1016/0749-596x(86)90021-5
- Kemmerer, D. (2012). The Cross-Linguistic Prevalence of SOV and SVO Word Orders reflects the sequential and hierarchical representation of action in broca’s area. *Lang. Linguist. Compass* 6, 50–66. doi: 10.1002/lnc3.322
- Kolomackiy, D. I. (2009). *Distribution of Russian Passive Forms: A Corpus Investigation*. Ph.D. dissertation, Moscow State University, Moscow.
- Maass, A., and Russo, A. (2003). Directional bias in the mental representation of spatial events: nature or culture? *Psychol. Sci.* 14, 296–301. doi: 10.1111/1467-9280.14421
- Mahowald, K., James, A., Futrell, R., and Gibson, E. (2016). A meta-analysis of syntactic priming in language production. *J. Mem. Lang.* 91, 5–27. doi: 10.1016/j.jml.2016.03.009
- Montag, J. L., and MacDonald, M. C. (2014). Visual salience modulates structure choice in relative clause production. *Lang. speech* 57, 163–180. doi: 10.1177/0023830913495656
- Myachykov, A., Garrod, S., and Scheepers, C. (2012a). Determinants of structural choice in visually situated sentence production. *Acta psychol.* 141, 304–315. doi: 10.1016/j.actpsy.2012.09.006
- Myachykov, A., Thompson, D., Scheepers, C., and Garrod, S. (2012b). Referential and visual cues to structural choice in visually situated sentence production. *Front. Psychol.* 2:396. doi: 10.3389/fpsyg.2011.00396
- Myachykov, A., Garrod, S., and Scheepers, C. (2018a). Attention and memory play different roles in syntactic choice during sentence production. *Discourse Process.* 55, 218–229. doi: 10.1080/0163853x.2017.1330044
- Myachykov, A., Pokhoday, M., and Tomlin, R. S. (2018b). “Attention and structural choice in sentence production,” in *Oxford Handbook of Psycholinguistics*, eds S.-A. Rueschemeyer and G. Gaskell (Oxford: Oxford University Press), 529–547.
- Myachykov, A., Posner, M. I., and Tomlin, R. S. (2007). A parallel interface for language and cognition in sentence production: theory, method, and experimental evidence. *Linguist. Rev.* 24, 457–474.
- Myachykov, A., Thompson, D., Scheepers, C., and Garrod, S. (2011). Visual attention and structural choice in sentence production across languages. *Lang. Linguist. Compass* 5, 95–107. doi: 10.1111/j.1749-818x.2010.00265.x
- Myachykov, A., and Tomlin, R. S. (2008). Perceptual priming and structural choice in Russian sentence production. *J. Cogn. Sci.* 6, 31–48. doi: 10.17791/jcs.2008.9.1.31
- Pokhoday, M., and Myachykov, A. (2018). Effects of auditory and visual cueing of attention on syntactic choice in sentence production. *Psychol. J. High. Econ.* 15, 209–221. doi: 10.1371/journal.pone.0195547
- Pokhoday, M., Scheepers, C., Shtyrov, Y., and Myachykov, A. (2018). Motor (but not auditory) attention affects syntactic choice. *PLoS One* 13:e0195547. doi: 10.1371/journal.pone.0195547
- Prat-Sala, M., and Branigan, H. P. (2000). Discourse constraints on syntactic processing in language production: a cross-linguistic study in English and Spanish. *J. Mem. Lang.* 42, 168–182. doi: 10.1006/jmla.1999.2668
- Rommers, J., Meyer, A. S., and Praamstra, P. (2017). Lateralized electrical brain activity reveals covert attention allocation during speaking. *Neuropsychologia* 95, 101–110. doi: 10.1016/j.neuropsychologia.2016.12.013
- Santiago, J., Román, A., Ouellet, M., Rodríguez, N., and Pérez-Azor, P. (2010). In hindsight, life flows from left to right. *Psychol. Res.* 74, 59–70. doi: 10.1007/s00426-008-0220-0
- Tomlin, R. S. (1995). “Focal attention, voice, and word order: an experimental, cross-linguistic study,” in *Word Order in Discourse*, eds P. Downing and M. Noonan (Philadelphia: John Benjamins), 517–554.
- Tversky, B. (2011). Visualizing thought. *Top. Cogn. Sci.* 3, 499–535. doi: 10.1111/j.1756-8765.2010.01113.x

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Pokhoday, Shtyrov and Myachykov. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

APPENDIX 1

Transitive events: hit, shoot, chase, touch, push, kick or “бить,” “стрелять,” “преследовать,” “трогать,” “толкать,” “пинать” in Russian, respectively. Referents: artist, chef, clown, cowboy, monk, nun, pirate, policeman, swimmer, dancer, professor, waitress, burglar, boxer, and soldier or “художник,” “повар,” “клоун,” “ковбой,” “монах,” “монашка,” “пират,” “полицейский,” “пловец,” “балерина,” “профессор,” “официантка,” “вор,” “боксер,” “солдат,” in Russian, respectively.



Speed-Accuracy Tradeoffs in Brain and Behavior: Testing the Independence of P300 and N400 Related Processes in Behavioral Responses to Sentence Categorization

Phillip M. Alday^{1*†} and Franziska Kretzschmar^{2,3†}

¹Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands, ²CRC 1252 "Prominence in Language", University of Cologne, Cologne, Germany, ³Institute of German Language and Literature I, University of Cologne, Cologne, Germany

OPEN ACCESS

Edited by:

Beatriz Martin-Luengo,
National Research University Higher
School of Economics, Russia

Reviewed by:

M. Van Hulle,
KU Leuven, Belgium
Yun Wen,
Aix-Marseille Université, France
Siri-Maria Kamp,
University of Trier, Germany

*Correspondence:

Phillip M. Alday
phillip.alday@mpi.nl

[†]These authors have contributed
equally to this work

Received: 28 February 2019

Accepted: 05 August 2019

Published: 27 August 2019

Citation:

Alday PM and Kretzschmar F
(2019) Speed-Accuracy Tradeoffs in
Brain and Behavior: Testing the
Independence of P300 and
N400 Related Processes in
Behavioral Responses to Sentence
Categorization.
Front. Hum. Neurosci. 13:285.
doi: 10.3389/fnhum.2019.00285

Although the N400 was originally discovered in a paradigm designed to elicit a P300 (Kutas and Hillyard, 1980), its relationship with the P300 and how both overlapping event-related potentials (ERPs) determine behavioral profiles is still elusive. Here we conducted an ERP ($N = 20$) and a multiple-response speed-accuracy tradeoff (SAT) experiment ($N = 16$) on distinct participant samples using an antonym paradigm (*The opposite of black is white/nice/yellow* with acceptability judgment). We hypothesized that SAT profiles incorporate processes of task-related decision-making (P300) and stimulus-related expectation violation (N400). We replicated previous ERP results (Roehm et al., 2007): in the correct condition (*white*), the expected target elicits a P300, while both expectation violations engender an N400 [reduced for related (*yellow*) vs. unrelated targets (*nice*)]. Using multivariate Bayesian mixed-effects models, we modeled the P300 and N400 responses simultaneously and found that correlation between residuals and subject-level random effects of each response window was minimal, suggesting that the components are largely independent. For the SAT data, we found that antonyms and unrelated targets had a similar slope (rate of increase in accuracy over time) and an asymptote at ceiling, while related targets showed both a lower slope and a lower asymptote, reaching only approximately 80% accuracy. Using a GLMM-based approach (Davidson and Martin, 2013), we modeled these dynamics using response time and condition as predictors. Replacing the predictor for condition with the averaged P300 and N400 amplitudes from the ERP experiment, we achieved identical model performance. We then examined the piecewise contribution of the P300 and N400 amplitudes with partial effects (see Hohenstein and Kliegl, 2015). Unsurprisingly, the P300 amplitude was the strongest contributor to the SAT-curve in the antonym condition and the N400 was the strongest contributor in the unrelated condition. In brief, this is the first demonstration of how overlapping ERP responses in one sample of participants predict behavioral SAT profiles of another sample. The

P300 and N400 reflect two independent but interacting processes and the competition between these processes is reflected differently in behavioral parameters of speed and accuracy.

Keywords: N400, P300, mixed-effects modeling, SAT, sentence processing, predictive processing

INTRODUCTION

Human cognition can be conceived of as a dynamic, hierarchically organized system of decision-making or categorization that accumulates evidence for (alternative) categories as new incoming sensory information is processed across time, and translates the outcome of this categorization to appropriate action once a decision threshold has been reached (Gold and Shadlen, 2007; Kelly and O'Connell, 2015).

Language is no exception to this: linguistic categorization is a dynamic process in which evidence from stimulus properties from lower to higher linguistic levels is accumulated across time, shaped by both stimulus-induced (exogenous) processes as well as decision-related (endogenous) processes. Associating sounds to phonemes, phoneme sequences to words and words to larger sentences are (somewhat simplified) examples for how humans categorize spoken linguistic input to compute the meaning of an utterance and subsequently plan an appropriate response. Importantly, predictive processing has been identified as a major (endogenous) mechanism in language comprehension that facilitates linguistic categorization in terms of processing speed and accuracy, as predictable linguistic units are processed faster and comprehended with fewer errors than unpredictable ones.

Our motivation for the current study is the observation that a fairly high number of studies on word recognition in isolation or in context report mixed evidence for effects of semantic prediction and relatedness/priming when comparing electrophysiological signatures such as event-related potentials (ERPs) with behavioral measures such as error rates (ER) and reaction time (RT). We restrict ourselves to studies that investigated how words are categorized as belonging to a certain semantic category by focusing on N400 and P300 ERPs in response to contextual predictability and semantic relatedness/priming with various experimental tasks (i.e., acceptability judgment, semantic categorization or comprehension tasks). As we will outline in more detail below, these studies reported a mixture of converging (i.e., identical effect directions of increases/decreases in ERP amplitudes, RT and ER) and diverging effects of these variables in the electrophysiological and behavioral data, a pattern that eludes a fully systematic explanation. More specifically, we conjecture that contextual predictability and semantic relatedness may impact ERPs differently than behavioral measures and that this interaction is additionally modulated by methodical complications. That is, cross-method divergence results in part from two well-known complications, namely that N400 and P300 overlap in time and scalp topography despite their different cognitive functions, and that standard RT and/or ER measures rely on a single data point insensitive to the

dynamics of categorization. This makes it difficult to unify, across electrophysiological and behavioral measures, effects of contextual predictability and semantic relatedness in signatures of stimulus processing and categorization at the word or sentence level.

The present article aims at presenting a novel cross-method approach to address this issue, and thereby to increase the validity of cross-method inferences on brain-behavior links or the perception-action loop in language processing—i.e., the time-course from neuronal processing (perception and categorization) to behavioral output (action). We specifically propose that the above complications may be overcome with time-sensitive behavioral measures such as the speed-accuracy trade-off (SAT) paradigm (Wickelgren, 1977) replacing standard RT measures and capturing decision dynamics more precisely, and with cross-method statistical modeling using mixed-effects models.

The N400 and the P300 are probably among the most intensively used ERP components to study language processing in humans and it is therefore not surprising that the range of their functional definitions varies tremendously. The following is thus not meant as a review of the extensive N400 and P300 literature but is highly selective in focusing on ERP-behavior relationships. The N400 is a negative-going deflection in the scalp-recorded EEG that peaks about 400 ms after the onset of a meaningful stimulus, showing a posterior maximum (Kutas and Federmeier, 2000, 2011). In particular for word recognition, the N400 has been found in response to words embedded in word lists, sentences and stories as well as in all modalities of language input (e.g., Kutas et al., 1987; Holcomb and Neville, 1990; Federmeier and Kutas, 1999; Alday et al., 2017). N400 amplitude is sensitive to a range of (broadly defined) semantic variables such as lexical frequency, contextual predictability, semantic relatedness/association, lexicality or orthographic neighborhood density (Kutas and Federmeier, 2000, 2011; Laszlo and Federmeier, 2009), but has also been found for processing at the syntax-semantics interface (e.g., Haupt et al., 2008; Bornkessel-Schlesewsky et al., 2011; Bourguignon et al., 2012) and discourse (e.g., van Berkum et al., 1999; Burkhardt, 2006). Predictability, including semantic priming as a subtype, has been found in particular to reduce N400 amplitude (Kutas and Federmeier, 2000; Federmeier, 2007; Van Petten and Luka, 2012). Building on this, it has been posited that amplitude increases to unpredictable input reflect either varying pre-activation levels of the target word, prediction mismatches between bottom-up input and top-down predictions or the extent to which perceived input does not match with the current resonance state of semantic memory (see Lau et al., 2008; Kutas and Federmeier, 2011; Lotze et al., 2011; Rabovsky and McRae, 2014; Bornkessel-Schlesewsky and Schlesewsky, 2019). Thus,

leaving aside the heterogeneous implementations of the proposed N400 models, an assumption common to all these accounts of the N400 is that its amplitude reflects the relative efficiency in processing stimulus or word properties in relation to the preceding context.

Although the majority of N400 studies report that reductions of N400 amplitude converge with reduced RT and error rates (or vice versa), there is also a non-negligible number of studies reporting diverging effects of N400 amplitude and behavioral measures. Many of the latter studies have investigated the processing of words either pre-activated/predicted *via* (lexical-)semantic priming, contextual predictability or a combination of both. The specific kind of divergence differs across studies, depending on whether: (i) N400 and behavioral measures show incongruent effect directions across measures or incongruent effect sizes, particularly nil effects in one vs. the other measure (e.g., Holcomb and Kounios, 1990; Kounios and Holcomb, 1992; Holcomb, 1993; Chwilla et al., 2000; Kiefer, 2001; Rolke et al., 2001; Federmeier et al., 2010; Debrulle et al., 2013; differences with eye movements: Dimigen et al., 2011; Kretzschmar et al., 2015; Degno et al., 2019); or (ii) behavioral effects have reflexes in a biphasic pattern of N400 and (partly) overlapping positivity (e.g., Roehm et al., 2007; Bakker et al., 2015; Meade and Coch, 2017). For instance, in a study on lexical and semantic-priming effects on the processing of newly-learned vs. existing words, Bakker et al. (2015) found diverging effects of lexicality and semantic relatedness in response accuracy and ERPs elicited by target words in a word-list presentation. Specifically, the interaction between lexicality and semantic relatedness affected response accuracy such that error rates were higher for novel words related to their prime than unrelated ones, but not for existing words. RT, by contrast, showed only a main effect of semantic relatedness such that related targets were responded to faster, regardless of the type of input (novel word vs. existing word). The interaction between lexicality and semantic relatedness affected ERPs somewhat differently in that the N400 was sensitive to semantic relatedness only with existing words, exhibiting the typical amplitude reduction for related words. The posterior late positivity showed an enhanced amplitude for existing and novel words following related primes, although this was qualified by the time that had elapsed between the learning and the test session. Specifically, the posterior priming effect based on semantic relatedness was only found with novel words that could consolidate in long-term memory, while there was no difference with more recently acquired novel words. Thus, online processing effects reflected in the N400 did not show up in behavior, while the late positivity showed an interaction only partly compatible with RT. While the correlation of behavioral and ERP data was not central to the research reported in Bakker et al. (2015), the authors suggested that component overlap of N400 and the late positivity may account for the lack of a priming effect for novel words in the N400 time window.

Indeed, component overlap seems to be a plausible explanation given an earlier finding that, with increasing strength of semantic relatedness and contextual predictability, ERPs in the N400 time window become more positive, resulting

in clearly visible P300 peaks for strongly related targets that can be actively predicted. This pattern was first reported in Kutas and Hillyard (1980) who showed that when context information and semantic relatedness converge to allow only one or a few candidates to felicitously end a sentence, N400 amplitude reduction seems to be overlaid with a P300. In other words, with high contextual constraint and a cloze probability of (nearly) 1 for the target, electrophysiological data are equivocal as to the ERP component driving amplitude modulations in the N400/P300 time window.

This pattern has been confirmed in a handful of ERP studies using the antonym paradigm (Bentin, 1987; Kutas and Iragui, 1998; Roehm et al., 2007; Federmeier et al., 2010) that provides strong semantic relatedness as well as high contextual predictability. Because antonyms are the logical endpoints on an opposition scale, antonym word pairs strongly prime each other. This effect can be strengthened with a sentence context such as *x is the opposite of y* or by using an experimental task that requires participants to think of or judge the antonymy relation between words, thereby increasing target cloze probability to nearly 1 (see Bentin, 1987; Roehm et al., 2007). Thus, from among the range of possible cloze probability values that a predictable target can have, the antonym paradigm picks up those with near-perfect cloze probability, yielding an almost binary distribution for predictable vs. unpredictable targets. Strikingly, even though this design revealed distinct P300 effects for expected antonyms and N400 amplitude increases for unpredicted non-antonyms across studies, the behavioral patterns do not converge with the ERPs. While some found that RT and error rates show facilitative effects for antonyms (Bentin, 1987), others found that non-antonym conditions fare better than antonyms behaviorally (Roehm et al., 2007; Federmeier et al., 2010). The lack of the typical behavioral priming effect for antonyms (i.e., reduced RT or error rates, see Neely, 1991) in some experiments is especially striking given that the ERP pattern is rather stable across studies.

This latter dissociation of P300 and behavioral measures in the antonym paradigm is also intriguing insofar as the P300/P3b, a domain-general positive-going potential that peaks about 250–500 ms after target onset and exhibits a posterior maximum (Polich, 2007), has been found to be sensitive to stimulus categorization and predictability and to show positive correlations with behavior. In particular, the P300 is elicited by motivationally significant target stimuli, especially those relevant for task performance (see reviews in Johnson, 1986; Nieuwenhuis et al., 2005, 2011; Polich, 2007). It has been linked to evidence accumulation for categorization, that is its amplitude is enhanced the more evidence from stimulus properties has been accumulated in order to make a decision on the stimulus category (O'Connell et al., 2012; Kelly and O'Connell, 2015; Twomey et al., 2015). As such it shows correlations with both stimulus-locked and response-locked brain activity (Verleger et al., 2005). More specifically, several studies have reported positive correlations between P300 latency and RTs (see review in Nieuwenhuis et al., 2005; for an example from language processing, see Sassenhagen and Bornkessel-Schlesewsky, 2015), as long as participants are instructed to emphasize response accuracy over speed (Kutas et al., 1977; but see Pfefferbaum et al., 1983).

In language processing, P300 latency varies with the absence or presence of a prediction match, especially when the target word is crucial to perform a categorization task with a binary choice (e.g., acceptability, sentence verification). For example, the P300 peaks earlier for the detection of a preferred (i.e., predicted) constellation than for a dispreferred or unpredicted one at various linguistic levels (see Haupt et al., 2008; Kretzschmar, 2010; Bornkessel-Schlesewsky et al., 2015; Graf et al., 2017). For instance, Graf et al. (2017) found that for grammatically correct vs. incorrect auxiliary choice in German sentences, P300 and acceptability judgments converged with grammatical auxiliary selection showing earlier P300 and higher acceptability ratings compared to ungrammatical selection. Similarly, in Roehm et al.'s (2007) study mentioned above, the P300 in response to predicted antonyms—the single possible sentence completion—peaked earlier than the P300 to unpredictable non-antonyms. Yet, when relevant stimulus properties conflict with one another and there is thus lower decision certainty during categorization, P300 amplitude is diminished. This is evidenced by some of the abovementioned studies investigating semantic relatedness. For instance, the P300 to non-antonyms in Roehm et al.'s (2007) study has a smaller amplitude when the non-antonym is semantically related to the predicted antonym compared to when it is unrelated (see “Experiment 1: Antonym Processing and ERPs” section below). Akin to what Bakker et al. (2015) reported for semantic priming for novel word meanings with a short consolidation time, P300 amplitude decreased for semantically related target words in Roehm et al.'s (2007) study. Importantly, however, behavioral data failed to converge with the ERP pattern, as antonyms did not show faster RT or higher accuracy than the other conditions.

In summary, both the N400 and the P300 appear to be sensitive to predictability during linguistic categorization: N400 amplitude and P300 latency each signal the presence or absence of a prediction match during target categorization, while semantic relatedness reduces the amplitude of both ERPs. Importantly, this pattern converges with proposals that the N400 indexes the processing of stimulus properties relevant for categorization (Bornkessel-Schlesewsky and Schlewsky, 2019, including linguistic fit), while the P300 indexes the dynamics of the categorization process itself (Twomey et al., 2015). Hence, N400 and P300-related processes depend on the same input, but reflect partly independent cognitive operations. A cognitive interpretation in terms of processing efficiency, however, is elusive as behavioral patterns (facilitation vs. inhibition) diverge.

Now, while aligning ERP patterns with behavioral patterns descriptively *via* inspection of their respective effect directions and sizes is not uncommon, it clearly suffers from two methodological challenges, summarized in (i) and (ii) below:

- (i) RT and accuracy are often measured with a single button press with substantial delay, i.e., seconds after the critical target engendering the ERP effect of interest. Standard RT measures thereby lack time-sensitive information about the development of the behavioral response or processing dynamics and reflect the unweighted sum of several online

processes. Inferences associating behavioral data to brain activity are thus difficult to draw. Related to this, standard RT measures conflate the likelihood of retrieving the correct information from memory with the likelihood to retrieve some representation faster than others (see McElree, 2006). Specifically, participants may trade speed for accuracy (i.e., give faster responses with a higher ER) or vice versa. Thus, any comparison between ERPs and behavior is complicated by the unidimensional nature of standard RT measures. This seems especially disadvantageous in cases as described above, where two distinct ERP components may index the categorization of stimulus properties and its associated time-course.

- (ii) N400 and P300 overlap in time and scalp topography. Thus, effects ascribed to either of the two components may also stem from processes related to the respective other component. That is, amplitude modulations in a given component under study may be the result of offsets introduced by an adjacent component (additive component overlap), reflective of modulations within a given component or a mixture of the two (multiplicative component overlap). This may interfere with the standard statistical analysis of ERPs, in which the two components are often investigated with voltage information from one and the same time window. From this perspective, where two components collapse towards a unidimensional voltage measure, inferences from electrophysiological to behavioral data are difficult to draw.

For the first issue, we propose that the SAT paradigm is better suited than standard RT measures to discover the time-course of decision-making during sentence categorization. The SAT method measures participants' binary decisions at varying latencies after the onset of the critical stimulus, thus capturing the development of categorization when information consolidates over time. In addition, with the SAT paradigm, categorization speed and accuracy can be dissociated analytically, as decision-making is reflected in three independent response parameters: asymptote, rate and intercept (Wickelgren, 1977). Response accuracy (measured in d' units) is reflected in the asymptote parameter. Speed parameters indicate when participants depart from chance level (intercept) and how quickly they achieve asymptotic performance (rate), i.e., their final decision state. The SAT paradigm may, therefore, allow for a more fine-grained comparison of ERPs and behavioral measures of processing efficiency because both data types capture some dimension of processing dynamics.

The second issue is more difficult to address in the presence of a biphasic ERP pattern. However, by applying modern statistical methods one can investigate the independence of the N400 and P300 signals. In using the antonym paradigm, the strong theory-based prediction of a P300 for a single possible completion and an N400 for violations of that prediction as well as the use of single-trial analyses incorporating both subject and item variation excludes the possibility that this biphasic pattern is artifactual (see Tanner et al., 2015 for filter artifacts, Tanner and Van Hell, 2014 for misleading grand averages in the case of interindividual differences). Joint modeling of both components

in the biphasic response, either through careful selection of covariates or through multivariate models, allows for adjusting for the influence of each component and modeling their covariance, respectively. For introducing our novel modeling approach and keeping model complexity reasonable, we focus on temporal overlap of the N400 and P300 occurring in the largely overlapping time windows (approximately 250–500 ms post target onset), as the ERP methodology in sentence and word processing is still more often used to make inferences based on the temporal dimension (i.e., when information is processed in the brain), rather than on an integrated spatiotemporal profile (but see Nieuwland et al., 2019). Hence, we will both disregard topographic overlap between the two components (although we note that this may be a useful extension of the approach) and the late positivity following the N400 for disconfirmed predictions, as, currently, it is not settled whether this is one component or several depending on topographical distribution (see Van Petten and Luka, 2012; Leckey and Federmeier, 2019).

We collected ERP and behavioral SAT data in two separate experiments to illustrate the feasibility of our proposal sketched above. Experiment 1 using ERPs serves as a replication of previous studies investigating categorization of predictable target words in sentences and Experiment 2 is complementary to standard RT measures accompanying ERP recordings. In both experiments, we used the antonym paradigm as presented in Roehm et al. (2007) and asked participants to judge sentences for acceptability on a binary (yes/no) scale.

EXPERIMENT 1: ANTONYM PROCESSING AND ERPs

Experiment 1 serves as a replication of the first experiment reported in Roehm et al. (2007). Roehm et al. (2007) investigated the comprehension of antonym pairs in a strongly constraining sentence “*x* is the opposite of *y*,” with *x* being the prime and *y* the target antonym (see example 1), where participants were asked to verify the antonymy relation between prime and target. The prime-target word pair is related *via* an antonymy relation and target predictability additionally strengthened *via* the sentence fragment occurring in between the two antonyms. The antonym pairs (example 1a) were contrasted with two types of violation, semantically related non-antonym targets (example 1b) and semantically unrelated non-antonyms (example 1c). This paradigm essentially contrasts the two variables predictability and semantic relatedness. In terms of predictability, only the antonym target is predictable from context, whereas both non-antonym endings are equally unexpected (see Roehm et al., 2007 for details about stimuli norming). Regarding semantic relatedness, related non-antonyms belong to the same semantic field or category as the expected antonym, whereas unrelated non-antonyms do not (see Löbner, 2013). Hence, semantic relatedness can be equated with semantic priming *via* an automatic spread of activation in long-term memory (see Collins and Loftus, 1975), while sentence contexts pushed predictions about what word can plausibly and truthfully end the sentence.

(1) Example sentences of the antonym paradigm employed in Experiment 1 by Roehm et al. (2007; target words are underlined)

- a. Black is the opposite of white.
- b. Black is the opposite of yellow.
- c. Black is the opposite of nice.

Roehm et al. (2007) found that strongly predicted antonyms, such as *white* in example (1a), engendered a P300 between 240 and 440 ms after target onset, which overlapped with the N400 that showed increased amplitudes for the two non-antonym conditions. The N400 effect was less pronounced for related non-antonyms from the same semantic category as the antonym (example 1b) vs. unrelated ones (example 1c). Additionally, N400 effects to non-antonyms were followed by a late positivity, which was stronger for unrelated non-antonyms than related non-antonyms at posterior electrode sites. These ERP effects are summarized in the top two rows of **Table 1**.

Although the antonym paradigm as described above includes a binary contrast between perfectly predictable targets and unpredictable violations, the ERP findings largely converge with previous studies which also manipulated target predictability and semantic relatedness. P300 responses to strongly predictable target words with near-perfect cloze probability (i.e., single possible completions, which is also the case for antonyms in sentence context), have been reported for word-list and sentence processing in English (Kutas and Hillyard, 1980; Bentin, 1987; Kutas and Iragui, 1998; Federmeier et al., 2010). Data from studies employing a broader range of cloze probability scores further support the pattern obtained in Roehm et al.’s (2007) experiment. P300 amplitude reductions as a consequence of semantic relatedness between target and prime have been previously found in a word-list experiment (Bakker et al., 2015). N400 amplitude increases to prediction violations and amplitude reductions due to semantic relatedness or category membership were reported for unexpected or unprimed words other than antonyms (e.g., Federmeier and Kutas, 1999; Bakker et al., 2015; Meade and Coch, 2017).

Overall, this pattern of results support the above considerations of how semantic relatedness/priming and predictability distinguish the three critical conditions in Roehm et al.’s (2007) design, and of how N400 and P300 ERPs

TABLE 1 | Summary of significant differences in Experiment 1 by Roehm et al. (2007).

ERP components and time windows analyzed	Effects
N400/P300 time window (240–440 ms)	antonym (P300) < related (N400) < unrelated (N400)
Late positivity time window (500–750 ms)	global distribution: antonym < related antonym < unrelated posterior distribution: related < unrelated
Error rate	unrelated < antonym < related
Response time	unrelated < antonym < related

Note: “*a* < *b*” means a significantly less negative amplitude (hence an increased P300 or reduced N400), lower error rate or shorter response time for *a* vs. *b*.

may index different aspects of linguistic categorization. The P300 indexes stimulus categorization and emerges within the N400 time window for prediction matches, especially when predictability and semantic relatedness converge to single out the expected target, here the second antonym word. In the case of prediction mismatches, P300 peak latency follows the N400 and its amplitude is reduced when semantic relatedness interferes with categorizing the stimulus as an unexpected non-antonym. The N400, in turn, overlays the P300 component when unpredicted stimulus features need to be processed. It shows facilitative effects of semantic relatedness for prediction mismatches, as priming facilitates the processing of stimulus features due to spreading activation and this is independent of the ensuing categorization.

Yet, the behavioral data from the antonymy verification task in Roehm et al.'s (2007) first experiment showed a pattern that is difficult to integrate with the above functional description of the ERP data, especially regarding the P300. For both ER and RT, unrelated violations (example 1c) were judged fastest and most accurate, whereas related violations (example 1b) were slowest and most error-prone. Antonyms fell in between the two prediction violations. Hence, behavioral data do not show clear evidence for a behavioral advantage of predictability that would mirror the P300 to antonyms, whereas they indicate that semantic relatedness of unpredicted non-antonyms has a negative effect, similar to the amplitude reduction of the late P300 in response to related non-antonyms. Conversely, the data are not suggestive of a facilitative behavioral effect of semantic relatedness that would mirror the N400 effect.

Given that the current experiment is a rather direct replication attempt of Roehm et al.'s (2007) first experiment, we expect to replicate both the ERP and behavioral data patterns.

Methods

Participants

Twenty participants (14 females, mean age: 23.15 years, SD: 2.60) from the University of Cologne participated for payment (8€/hour) or course credit. All participants were monolingual native speakers of German and reported normal or corrected-to-normal vision and no history of psychological or neurological disorders. All were right-handed as assessed with an abridged German version of the Edinburgh handedness test (Oldfield, 1971). The protocol for ERP experiments conducted in the lab is approved by the Ethics Committee of the German Society of Linguistics (DGfS; #2016-09-160914). Participants gave written informed consent prior to their participation.

Materials

We used the same sentence stimuli as in Roehm et al. (2007) and made publicly available in Roehm (2004).

Apparatus and Procedure

EEG was recorded from 55 Ag/AgCl electrodes (ground: AFz; 10-10 system) fixed at the scalp by means of an elastic cap (Easycap GmbH, Herrsching, Germany). EOG was recorded from three additional pairs of electrodes placed at the outer canthus, supraorbital and infraorbital of each eye. The sampling rate was 500 Hz (BrainAmp DC, Brain Products, Gilching,

Germany). Data were referenced to the left mastoid for recording. Impedances were kept below 5 kOhm.

Before the experiment, participants were instructed to judge in an acceptability task whether the sentence is correct or not, and were given 10 practice trials to familiarize with the task. Note that we did not use the kind of antonym verification judgment employed in the original study, as this was less optimal for Experiment 2 (see "Apparatus and Procedure" section below). Participants were seated in a sound-attenuated booth, at a distance of approximately 100 cm from a 24-inch monitor. Sentences were displayed centered on the screen and in black font (Verdana, 28 pt) against a light-gray background. Rapid serial visual presentation (RSVP) closely followed the specifications given for Roehm et al.'s (2007) first experiment [with the exception of the inter-trial interval (ITI)]. Each trial began with the presentation of a fixation star, presented for 2,000 ms, to focus participants' attention to the upcoming sentence. Sentences were then presented word by word, with 350 ms per word and 200 ms interstimulus interval (ISI). After the sentence-final target word, a blank screen was presented for 650 ms and then replaced with question marks indicating that participants could now give their judgment with one of two buttons on a game pad. Maximum response time was 3,000 ms. The ITI was 2,000 ms (vs. 2,250 ms in the original study). Assignment of response buttons (correct vs. incorrect) to the right and left hand was counterbalanced across participants.

Items were presented in four lists, each containing 80 sets of antonym sentences and 40 sets in each of the two non-antonym conditions. Participants were randomly assigned to one of the lists, which were presented in one of two pseudorandomized orders.

Analysis and Results

EEG data were processed with MNE-Python 0.17.1 (Gramfort et al., 2013). Data were re-referenced to linked mastoids offline and bandpass filtered from 0.1 to 30 Hz (bandpass edge, hamming-windowed FIR, with zero-phase achieved *via* compensation for the group delay). Bipolar horizontal and vertical EOG were computed, and the very most anterior (AFx), posterior (Px) and temporal electrodes (TPx) data were excluded from further analysis. The continuous EEG was then divided into epochs extending from 200 ms before onset of the critical word until 1,200 ms after onset. Trials where the peak-to-peak voltage difference exceeded 150 μ V in the EEG or 250 μ V in the bipolar EOG were excluded from further analysis. Additionally, flat-line trials (where the peak-to-peak voltage in the EEG was less than 5 μ V) and trials where the absolute voltage exceeded 75 μ V were excluded. No baseline correction was performed as part of the preprocessing. However, the trial-wise mean voltage pre-stimulus interval (−200 to 0 ms) was used to baseline correct for plotting purposes and entered as a covariate into the statistical analyses (see Alday, 2017). The preprocessed EEG data along with analysis source code is available on the Open Science Framework (OSF; see "Data Availability Statement" below).

Subsequently, trials with an incorrect or timed-out behavioral response were also excluded (2%–5% of trials on average per condition). As this reflects ceiling performance, we did not

further analyze behavioral data from the EEG experiment. However, numerical values for both RT and accuracy rates are highly similar to the original data, as shown by grand means and standard errors: highest accuracy rates (0.98 ± 0.012) were obtained for the unrelated non-antonyms, followed by the antonym condition (0.95 ± 0.016). Related non-antonyms were judged with lowest accuracy (0.94 ± 0.012). RT to correctly answered trials confirmed this pattern, with fastest RT (in milliseconds) for unrelated non-antonyms (450 ± 37), slowest RT for related ones (550 ± 57), and antonyms falling in between the two (470 ± 31).

In total, 2,898 trials across 20 subjects remained for an average of 145 trials per participant (72 antonym, 36 related, 37 unrelated).

Figure 1 shows the grand-average response at Cz with 83% confidence intervals. Non-overlap of 83% confidence intervals corresponds to significance at the 5% level, or equivalently, the 95% confidence interval of the difference not crossing 0. As expected and observed in previous studies, we see a clear P300 for the antonym condition and a graded N400 for the related and unrelated violation conditions. As shown in the by-condition plots (**Figure 2**), the topographies of these components correspond to the typical centro-parietal characterization of the P300 and N400 components.

As the purpose of this study was not to examine the topography of well-characterized components, we restrict ourselves for simplicity and computational efficiency in the cross-method analysis to a centro-parietal region of interest (ROI) comprising 26 electrodes (C1, C2, C3, C4, C5, C6, Cz, CP1, CP2, CP3, CP4, CP5, CP6, CPz, P1, P2, P3, P4, P5, P6, P8, Pz, PO3, PO4, POz, Oz) that were least affected by artifacts across participants and trials, and that typically show maximum activity for the visually-evoked N400 effect (e.g., Johnson and Hamm, 2000) and P300 effect (e.g., Verleger et al., 2005), respectively. We used single trial mean voltage for the *a priori* chosen P300 (200–300 ms post-stimulus, as this more adequately captured P300 activity for antonyms, see Bentin, 1987; Roehm et al., 2007) and N400 (300–500 ms post-stimulus, see Kutas and Federmeier, 2000), and this was used as EEG measure in all analyses below. While the choice of component time windows reduces overlap, it does not eliminate it, if for no other reason than a larger P300 serves as an offset for a subsequent N400 component.

We analyzed these single-trial data with linear mixed-effects models using lme4 (v1.1-20, Bates et al., 2015b), with fixed effects for the mean voltage in the baseline window (see above; Alday, 2017) and condition as well as their interaction. All EEG measures were transformed to the standard deviation scale, and condition was sequential difference coded such that the contrasts related > antonym and unrelated > related are directly represented in the coefficients.

Random effects consisted of by-item intercepts and by-subject intercepts and slopes for condition. This models random variation in the lexical material as well as between-subject differences in the overall and by-condition EEG response. While this random-effect structure is not maximal in the sense of Barr et al. (2013), the data do not support a more complex structure and we do not expect additional variation along the

omitted dimensions (see Bates et al., 2015a; Matuschek et al., 2017). Moreover, for the present study, where model comparison is more important than significance, any potential issues with anti-conservative significance of fixed-effects component are irrelevant.

Statistical analysis confirms the visual impressions that the present data replicate the findings of Roehm et al. (2007; see **Tables 2, 3**). In particular, we observe a graded response in both the N400 and P300 time windows, with the main effect for condition reflecting a significant difference between related and unrelated (the reference level) as well as antonym and unrelated.

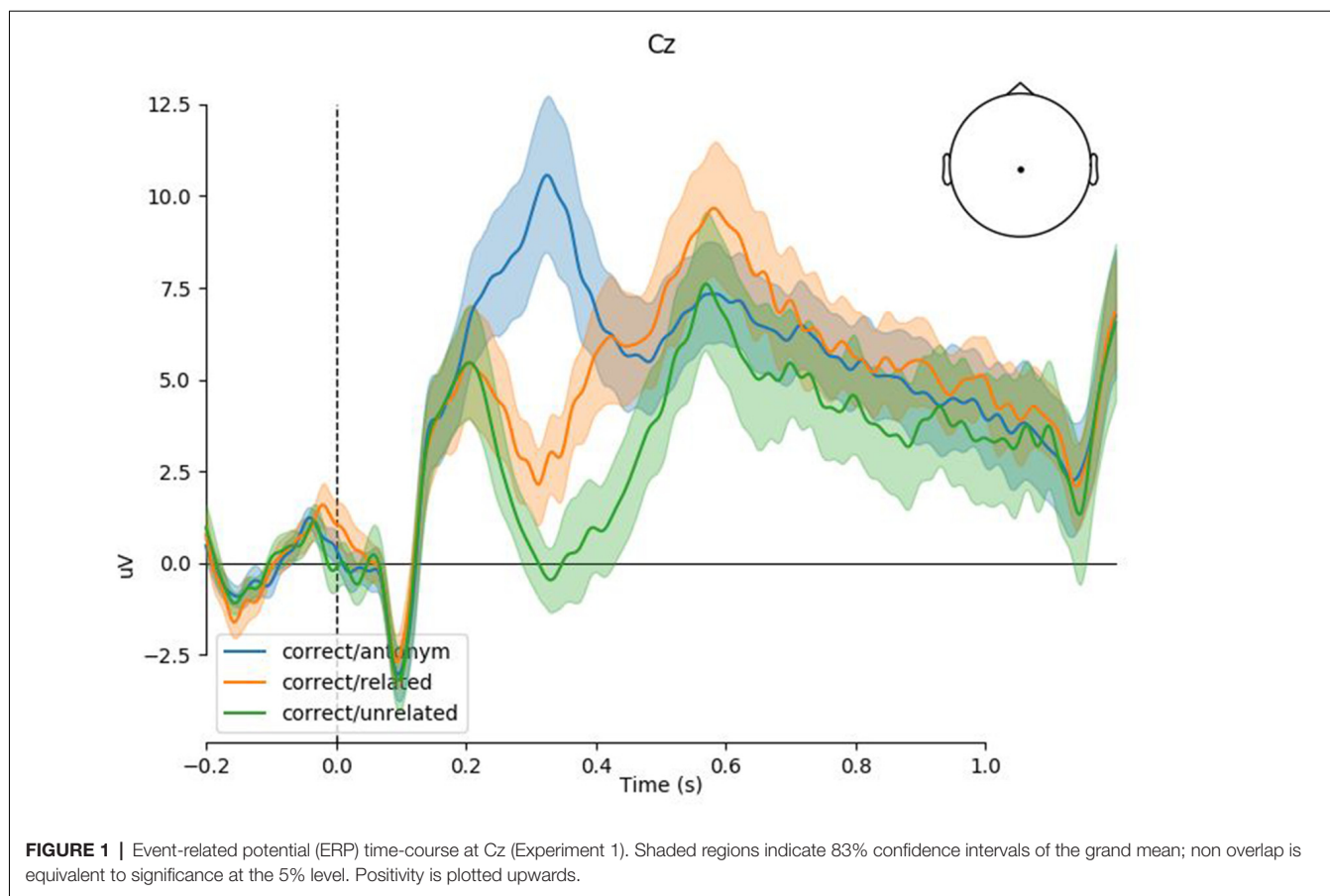
Discussion of Experiment 1

The current experiment aimed at replicating the findings from Experiment 1 in Roehm et al. (2007). In line with the original study, we find that the conditions elicit distinct ERP responses depending on target predictability and semantic relatedness. Between 200 and 300 ms post target onset, antonyms (*white*) engender a pronounced P300, while related non-antonyms (*yellow*) and unrelated non-antonyms (*nice*) both elicit an N400 effect between 300 and 500 ms post target onset. The N400 for unrelated non-antonyms was larger than the one for related non-antonyms. In addition, visual inspection suggested that the N400 in the two non-antonym conditions was followed by a late positivity, which was, however, less pronounced than the early P300 for antonyms.

EXPERIMENT 2: ANTONYM PROCESSING IN THE SPEED-ACCURACY TRADE-OFF PARADIGM

As discussed above, with standard behavioral measures of response time and accuracy, data interpretation can be complicated by the fact that response time and accuracy may vary in their relationship across participants and on a trial-to-trial basis. That is, participants may trade response speed for accuracy or vice versa, for instance when adapting their decision criterion to the experimental task at hand (see Kutas et al., 1977; Wickelgren, 1977).

In Experiment 2, we used the SAT paradigm (Wickelgren, 1977) that measures participant' response accuracy as a function of their response speed and that has been successfully employed in a number of previous investigations on various phenomena in sentence processing (e.g., McElree et al., 2003; Bornkessel et al., 2004; Martin and McElree, 2009; Bott et al., 2012). We adopted the SAT paradigm as it allows independent estimates of processing accuracy and dynamics. Participants give speeded binary acceptability judgments in response to short signal tones, presented at varying latencies from critical word onset. Individual d' scores are computed as a measure of sensitivity to stimulus properties and the development of response accuracy depending on time is described with three SAT parameters. Asymptote (λ) reflects the highest level of participants' accuracy. Response speed is reflected in two parameters: the intercept (δ) is the point when participants depart from chance level in giving accurate responses and the



rate (β) reflects the speed with which they reach their individual asymptotic performance. Thus, the categorization process can be described with multidimensional behavioral data (contrasting with standard RT measures).

We predict that the three conditions in the antonym paradigm should exhibit distinct SAT profiles. Recall that only antonym pairs are predictable, whereas the two non-antonym conditions are unpredictable from the preceding context. Related non-antonyms are distinct from unrelated ones by being semantically related to the correct and predicted antonym. Specifically, there are two possible general predictions based on whether: (a) predictability dominantly determines categorization or (b) whether predictability and semantic relatedness interactively determine decision. If only predictability matters for categorization, then decisions for antonyms should be more accurate and faster than the other two unexpected sentence endings. If, however, in addition to predictability semantic relatedness is taken into account for categorization, we expect a slightly different pattern. Specifically, semantic relatedness may be helpful in stimulus processing under the premise of spreading activation of the expected antonym to other category members (see Collins and Loftus, 1975; Kretzschmar et al., 2009). However, from the perspective of categorization, relatedness may likewise be conceived of as an intervening factor in deciding on whether, e.g., *yellow* is or is not an antonym to *white*. By definition,

category members share semantic features which makes their categorization less easy for related non-antonyms as they are less distinct from the expected antonym by means of shared features. Features shared between the expected target and a competitor (cue overload) has been shown to make other categorization at the sentence level (e.g., subject-verb agreement) harder, leading to lower accuracy and slower processing dynamics in the SAT curve (McElree et al., 2003; Johns et al., 2015). Thus, if semantic relatedness is indeed an intervening factor in categorization, related non-antonyms should show lower asymptote and slower processing dynamics compared to the other two conditions because it is more difficult to achieve a stable decision point. Antonyms and unrelated non-antonyms should reveal identical patterns from this perspective because decision can be reliably made due to a prediction match (i.e., identical feature set of expected target and perceived target) or a mismatch with unshared features (i.e., maximally distinct feature set for unrelated non-antonyms compared with the expected antonym).

Note that our predictions for differences in processing speed are somewhat speculative because previous results on the retrieval of semantic cues in sentence processing using the SAT method have provided mixed findings on differences in processing dynamics (e.g., McElree et al., 2003; Martin and McElree, 2009; Johns et al., 2015).

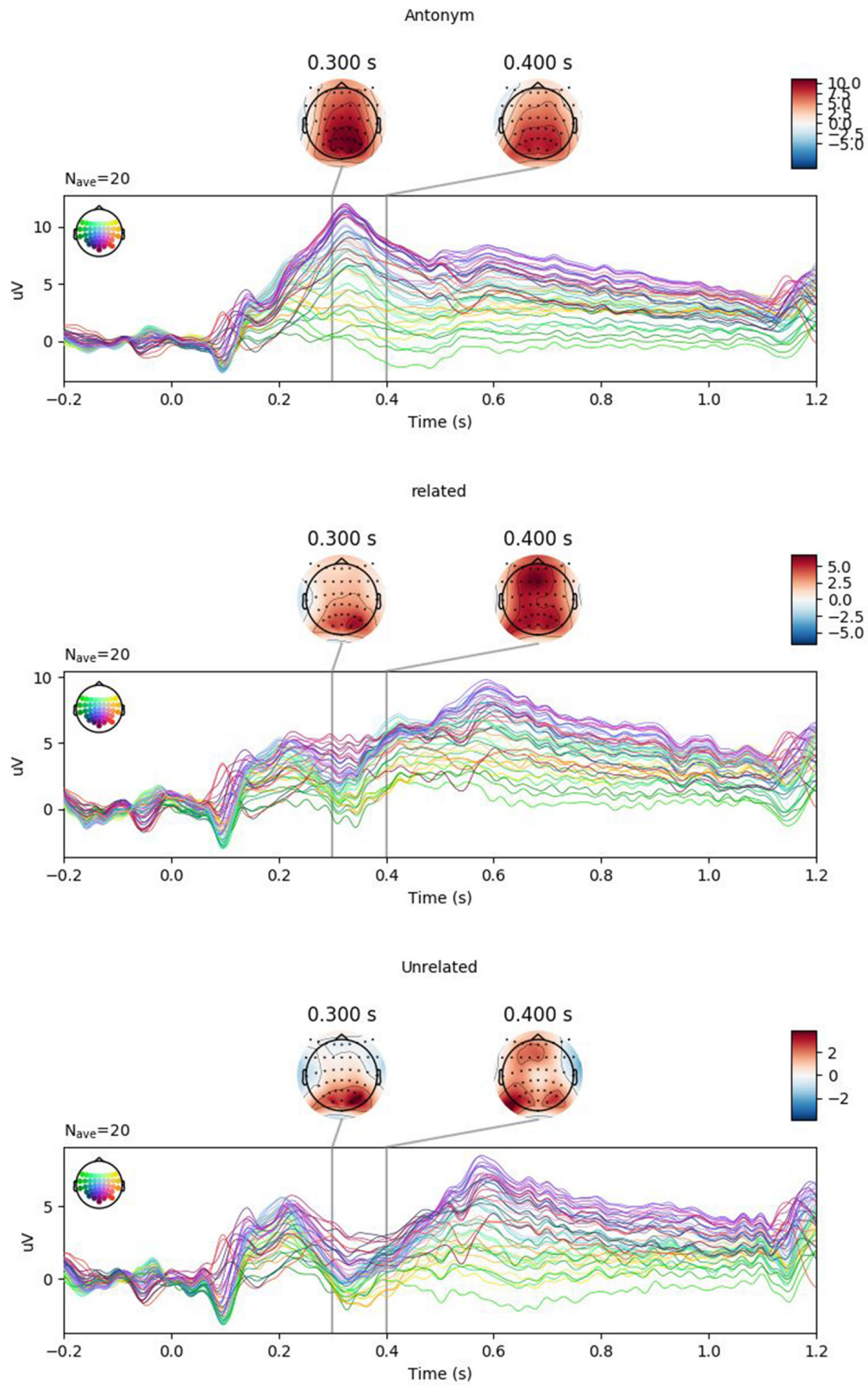


FIGURE 2 | Topographies of ERPs by condition (Experiment 1). Positivity is plotted upwards. Note the clear positivity peaking at 300 ms in the antonym condition (top) as well as the negativity around 400 ms in the related (middle) and unrelated (bottom) conditions.

TABLE 2 | Linear mixed effect model for the P300 time window (Experiment 1).

AIC	BIC	logLik	deviance	df.resid
7,743	7,827	-3,858	7,715	2,884
Scaled residuals:				
Min	1Q	Median	3Q	Max
-4.45	-0.64	-0.02	0.64	3.61
Random effects:				
Groups	Term	Std.Dev.	Corr	
item	(Intercept)	0.11		
subj	(Intercept)	0.25		
	related >	0.45	-0.16	
	antonym			
	unrelated >	0.10	-0.98	0.34
	related			
Residual		0.89		
Number of obs: 2,898, groups: item, 80; subj, 20.				
Fixed effects:				
	Estimate	Std. Error	t-value	
(Intercept)	-0.089	0.059	-1.5	
baseline	0.09	0.02	4.6	
related >	-0.47	0.11	-4.3	
antonym				
unrelated >	0.044	0.052	0.84	
related				
baseline:related >	0.11	0.045	2.4	
antonym				
baseline:unrelated >	-0.022	0.05	-0.45	
related				

The response is the trial-wise mean amplitude at a centro-parietal ROI in the time window 200–300 ms, the baseline is the trial-wise mean amplitude in the pre-stimulus window -200 to 0 ms. EEG measures are centered and scaled. Model estimated using maximum likelihood (i.e., REML = FALSE) and the bobyqa optimizer.

TABLE 3 | Linear mixed effect model for the N400 time window (Experiment 1).

AIC	BIC	logLik	deviance	df.resid
7,002	7,086	-3,487	6,974	2,884
Scaled residuals:				
Min	1Q	Median	3Q	Max
-3.29	-0.66	0.03	0.65	5.06
Random effects:				
Groups	Term	Std.Dev.	Corr	
item	(Intercept)	0.14		
subj	(Intercept)	0.30		
	related >	0.37	-0.41	
	antonym			
	unrelated >	0.12	0.40	0.68
	related			
Residual		0.78		
Number of obs: 2,898, groups: item, 80; subj, 20.				
Fixed effects:				
	Estimate	Std. Error	t-value	
(Intercept)	-0.14	0.071	-1.9	
baseline	-0.17	0.017	-9.7	
related >	-0.55	0.091	-6	
antonym				
unrelated >	-0.32	0.05	-6.4	
related				
baseline:related >	0.064	0.039	1.7	
antonym				
baseline:unrelated >	-0.0053	0.044	-0.12	
related				

The response is the trial-wise mean amplitude at a centro-parietal ROI in the time window 300–500 ms, the baseline is the trial-wise mean amplitude in the pre-stimulus window -200 to 0 ms. EEG measures are centered and scaled. Model estimated using maximum likelihood (i.e., REML = FALSE) and the bobyqa optimizer.

Methods

Participants

Sixteen participants (nine females, mean age: 24.44 years, SD: 2.61) from the Universities of Marburg and Mainz participated in Experiment 2. Participants were paid 7€/hour for their participation. None of them participated in Experiment 1. All participants were native speakers of German (15 monolingual, one bilingual) and reported normal or corrected-to-normal vision and no history of psychological or neurological disorders. Experiment 2 was not accompanied by an ethics vote but was conducted in line with national and institutional guidelines, as specified by the rules of the German Research Foundation (DFG). Specifically, behavioral non-invasive experiments with healthy young adults (between 18 and 65 years) do not require one as long as they pose no risk or physical/emotional burden to participants and as long as participants are debriefed after participation. See “Ethics Statement” for details. Participants gave written informed consent prior to their participation. One participant was excluded from analysis because of below-chance performance in response accuracy.

Materials

We selected 20 sets of items from the original 80 sets used in Experiment 1. The number of items was reduced in order to keep the number and length of experimental sessions at a reasonable size, as SAT experiments are typically conducted with many more filler items than ERP experiments. There were eight items with adjectival pairs and six items with verbal and nominal

pairs each. The order of prime and target words was reversed to meet methodical requirements of the SAT procedure: in order to obtain a useful estimate of processing speed, the critical target word needs to be lexically identical across conditions. By reversing prime and target words in the original item sets, we could achieve that (see example 2). Each item occurred in one of the three critical conditions (antonym, related and unrelated non-antonyms) and in a fourth repetition condition that was used for d' scaling.

(2) Example set of items in Experiment 2

- antonym condition: *Klein ist das Gegenteil von groß.* “Small is the opposite of big.”
- related non-antonym: *Dick ist das Gegenteil von groß.* “Thick is the opposite of big.”
- unrelated non-antonym: *Grün ist das Gegenteil von groß.* “Green is the opposite of big.”
- repetition: *Groß ist das Gegenteil von groß.* “Big is the opposite of big.”

With the acceptability task used here, the antonym condition is the only one requiring an “acceptable” (yes) response. There were 40 filler items with a comparable sentence beginning (“x is the y”) to reduce the saliency of the frame “x is the opposite of y”; 20 of them contained semantic or syntactic (gender, category) violations at various positions in the sentence, thus requiring an “unacceptable” (no) response. There were further 336 filler sentences of varying structures from other experiments, 184 of

which required an “unacceptable” (no) response. From the total of 464 sentences, 264 (57%) required an “unacceptable” (no) response, 200 (43%) an “acceptable” (yes) response¹. Items in the four critical conditions constituted 17% of all trials.

Apparatus and Procedure

Items were presented in black font (Monaco, size: 38 pt) on a white background, centered at the screen of a 21-inch monitor. Participants were instructed to read the sentences and to judge them for acceptability (yes/no) upon hearing a response signal. We did not use an antonym verification task as in the original study by Roehm et al. (2007) because this would have not worked for the various filler items.

We employed the multiple response-SAT paradigm (see Bornkessel et al., 2004; Martin and McElree, 2009). Fifteen response tones (2,000 Hz, 50 ms duration) followed each sentence, with the first two tones preceding the onset of the target word that provides the essential piece of information to judge acceptability. Participants had to give their response within 300 ms following each tone. Each trial began with a fixation star presented for 400 ms and an ISI of 1,000 ms. Next, participants saw which of the two response buttons (y and n on the keyboard) would serve as the default button for the responses in which they could not yet give a certain answer (see Bornkessel et al., 2004). Occurrence of the default buttons was equibalanced within and across conditions. Then, sentences were presented word-by-word at a fixed presentation rate of 300 ms/word and with an ISI of 100 ms. Before the onset of the sentence-final target word, the first two response tones were presented right after the offset of the pre-final word, and participants had to press the default button as a response within 300 ms following each of the two tones. Participants were instructed to switch to the y button for “acceptable” responses or the n button for “unacceptable”

¹A reviewer noted that in non-SAT lexical-decision and semantic-relatedness experiments, participants typically respond faster with positive/yes answers than negative/no answers and expressed concern that comparing yes and no responses is thus an unfair comparison. To address this concern, we note that the RT difference between the response polarity exactly follows the experimental manipulation, i.e., is not typically separable from the effect of condition. Moreover, this is line with current computational and psychological theory. Verifying a word is much faster than rejecting a nonword because the word can be accepted as soon as a match is found in the mental lexicon, while an exhaustive search is necessary for a nonword. Or in terms of activation: the baseline activation of a real word is much higher than a nonword and so processing is easier and faster. Second, in our case, the difference between yes and no responses is somewhat separable from the effect of condition because there are multiple conditions which require a no response (all but the actual antonym condition) and which nonetheless differ in their SAT curves. That is, unrelated non-antonyms are different from related ones in speed and accuracy, even though both require participants to give a No response in order to correctly perform the task. This suggests that the differences between conditions are at least in part due to the experimental manipulation and not the yes/no distinction. Of course, having multiple No conditions and only one Yes condition results in a lack of balance. In the overall experiment, this is eliminated *via* the filler sentences. For the analysis of the critical sentences, this is not problematic for the statistical methods used, as the dependent variable was encoded as accuracy and not response polarity. Finally, the nature of RT within the multiple response-SAT paradigm, where the response latencies are largely determined by the experimenter and not the participant, should preclude any such differences. Indeed, this is born out in the data, with no differences between positive and negative responses (see “Data Availability Statement” below for links to the data and scripts on OSF).

responses as soon as they could make a decision after seeing the target word on screen. The next trial began after an ITI of 1,500 ms.

The items were allocated to two lists; each list (containing 232 trials) was presented in eight blocks with short breaks in between. The first session additionally comprised a practice with 50 sentences unrelated to the experimental items, in which participants were trained to respond within 300 ms after tone onset. Participants took part in the sessions on two consecutive days.

Analysis and Results

Before analysis, the data from all participants were preprocessed to remove invalid data points. Due to recording bugs in presentation, some trials contained excessively long pauses before or during tone presentation. These trials were excluded from analysis (3.3% of trials), as were timed-out responses that did not occur within 300 ms after signal tone offset (less than five responses per condition on average across participants and latencies). The preprocessed SAT data along with analysis source code is available on the OSF (see “Data Availability Statement” below).

For an initial assessment of behavioral performance, accuracy was computed for each decision point during the response interval (per participant and condition), using d' as a sensitivity measure. Hits were defined as yes/“acceptable” responses to the antonym condition (example 2a) and no/“unacceptable” responses to the two non-antonym conditions (examples 2b, c). False alarms were defined as yes-responses to the repetition condition (example 2d). The resulting mean SAT curve is shown in **Figure 3**. In terms of percentage correct, the identity, unrelated and antonym conditions all reached ceiling (respective grand mean accuracies and standard errors at the final tone: 0.97 ± 0.027 , 0.96 ± 0.022 , 0.97 ± 0.008), while the related condition showed slightly worse but still high performance (0.83 ± 0.036), with the decreased performance perhaps reflecting interference and decision uncertainty (discussed more below).

In contrast to traditional SAT analysis using within-subject curve-fitting to a subject’s d' time-course with an exponential decay of error towards an asymptote, we used mixed-effects logistic regression with by-trial accuracy to model the SAT (see Davidson and Martin, 2013). This method has a couple of advantages for the present study: (1) we are not dependent on aggregation and can thus model item variance as well as trial-by-trial fluctuation in RT to each tone; and (2) we can model all subjects and their associated variance in a hierarchical fashion, allowing for partial pooling and shrinkage. This should yield more robust inferences. The overall time-courses for both methods are comparable, as seen in **Figures 3, 4**.

Fixed effects consisted of log-transformed total RT (tone latency + response time to that tone), condition and their interaction. Condition was sequential-difference coded in the same way as for the EEG data. Again, similar to the EEG data, random effects consisted of by-item intercepts and by-subject intercepts and slopes for log RT.

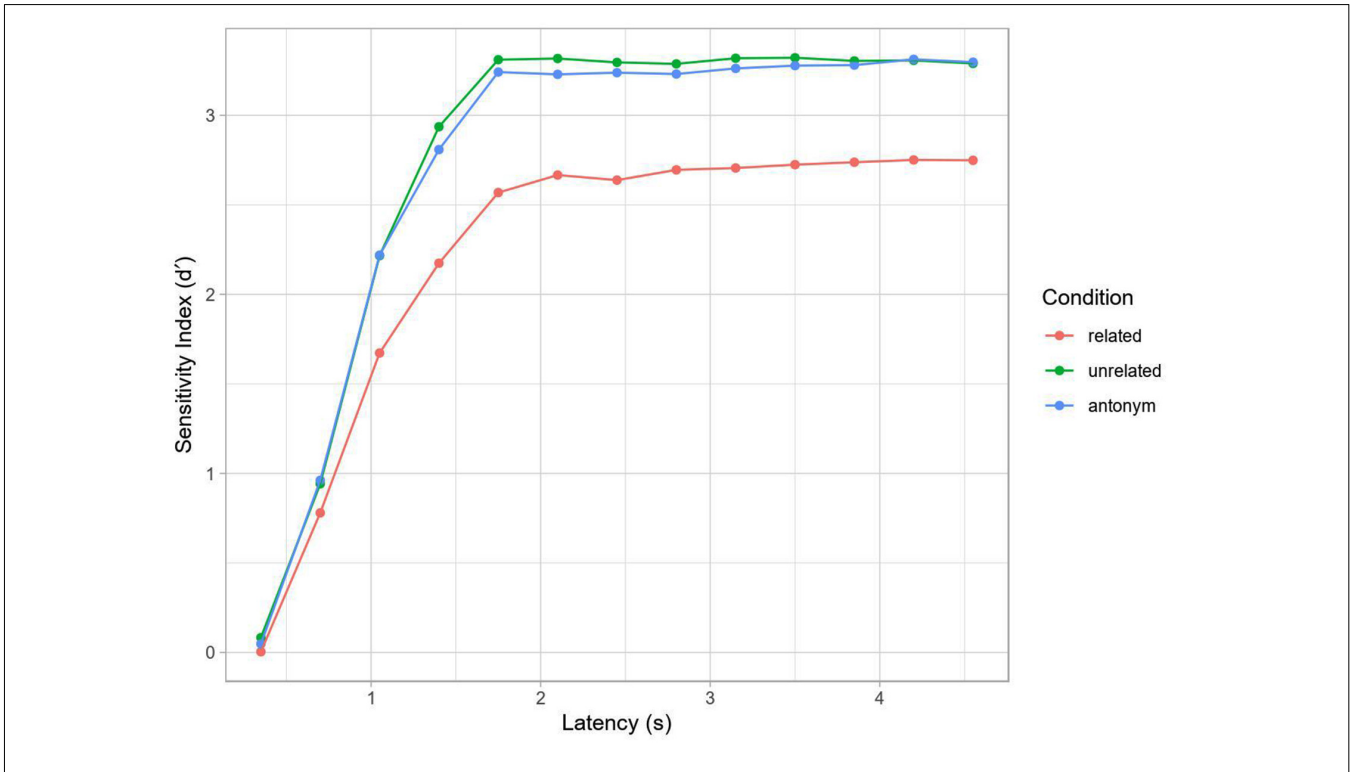


FIGURE 3 | Speed-accuracy tradeoff (SAT) by condition (Experiment 2). Curve computed on grand average data. Note the lower asymptotic performance in the related condition, but otherwise similar dynamics.

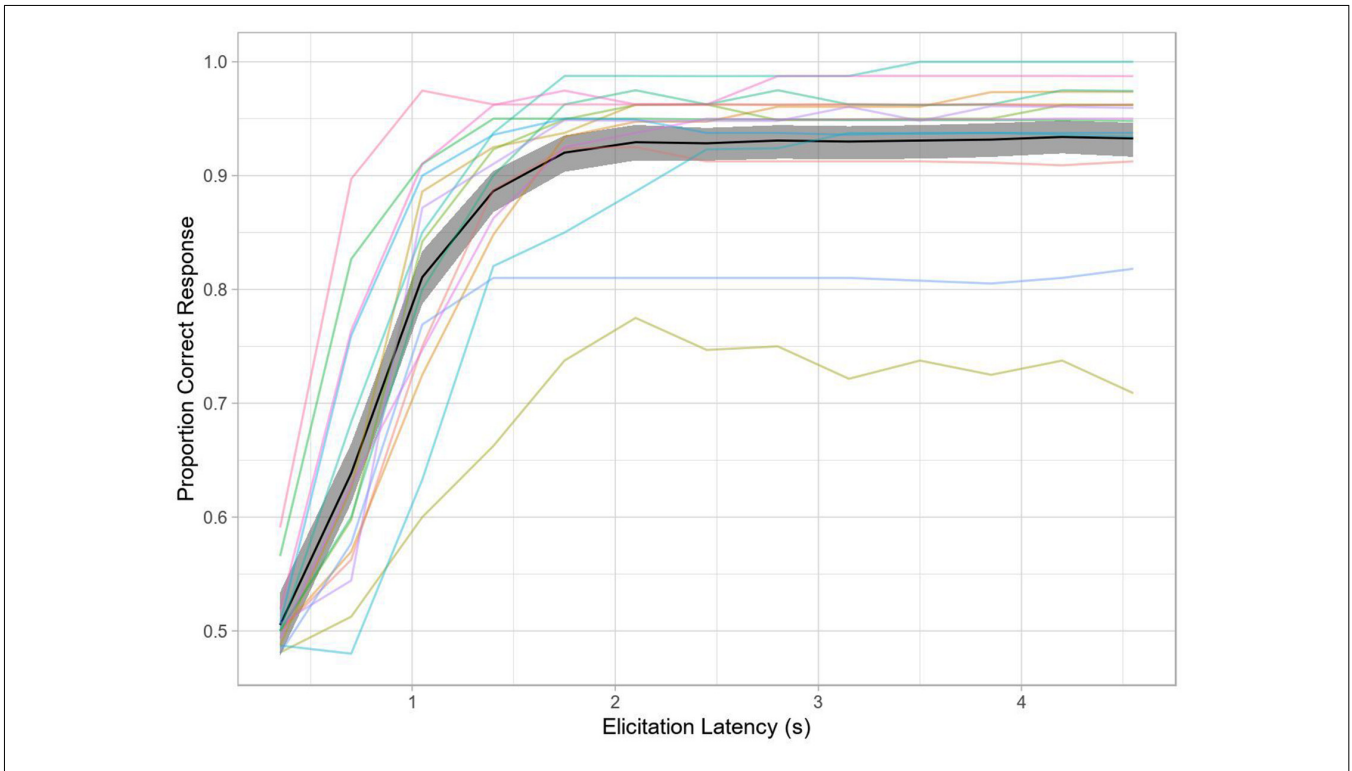


FIGURE 4 | Accuracy over time (Experiment 2). The black line represents grand average accuracy across all conditions, with the shaded region indicating the 95% bootstrapped confidence interval of the grand mean. The colored lines indicate single-subject performance.

TABLE 4 | Generalized linear mixed effect model for the speed-accuracy tradeoff (SAT) data (Experiment 2) based on condition.

AIC	BIC	logLik	deviance	df.resid	
7,293	7,367	-3,637	7,273	11,484	
Scaled residuals:					
Min	1Q	Median	3Q	Max	
-15.05	0.09	0.2	0.37	2.95	
Random effects:					
Groups	Term	Std.Dev.	Corr		
item	(Intercept)	0.65			
subj	(Intercept)	2.81			
	logRT	0.45	-0.996		
Number of obs: 11,494, groups: item, 20; subj, 15.					
Fixed effects:					
	Estimate	Std. Error	z-value	Pr (> z)	
(Intercept)	-10	0.81	-12	1.3e-35	***
related > antonym	6.4	0.73	8.8	1.4e-18	***
unrelated > related	-5.8	0.73	-13	1.2e-15	***
logRT	1.7	0.13	13	1.6e-39	***
related > antonym:logRT	-1.1	0.1	-10	1.5e-25	***
unrelated > related:logRT	0.98	0.1	9.7	3.3e-22	***

Dependent variable is the response accuracy, correspondingly the model family is binomial with a logit link. RT is the total reaction time, i.e., the response tone latency plus the reaction time to that tone. Model fit by maximum likelihood using the Laplace approximation and the bobyqa optimizer.

Although performance for later latencies was generally near ceiling, the related condition showed a significantly lower asymptotic performance than the other conditions (as shown in the combination of the intercept, and interaction effects for condition, **Tables 4, 5**, see also **Figure 3**) and a slower ramp-up (as shown in the interaction effects for condition and log RT). This is comparable to a difference in the asymptote and rate parameters in traditional SAT analysis.

Discussion of Experiment 2

Experiment 2 is, to the best of our knowledge, the first experiment to investigate antonymy processing in the SAT paradigm. We hypothesized that speed and accuracy parameters are differentially influenced by the conditions, either due to predictability alone or due to an interaction of predictability and semantic relatedness. We found significant differences between conditions both in asymptotic performance and in processing dynamics (reflected in rate). Related non-antonyms were rated less accurately and at a slower rate than the other two conditions that did not differ from each other. The results thus suggest that predictability alone does not influence processing accuracy and speed in the antonym paradigm, because antonyms did not differ from both non-antonym conditions. Rather, semantic relatedness and predictability interacted such that relatedness made the evaluation of a target word as a prediction mismatch more difficult.

These findings support and refine previous behavioral data obtained in the antonym paradigm. The SAT data confirm that related non-antonyms are in fact more difficult to judge, as reflected both in RT and accuracy. This lends further support to our hypothesis that semantic relatedness interferes with categorization in that only related non-antonyms contain information that impede an unequivocal categorization. At the same time, the SAT data do not reveal significant differences between antonyms and unrelated non-antonyms as previously found with standard RT measures. This can be explained with

TABLE 5 | Comparison of slopes across conditions in the SAT model.

Contrast	Estimate	SE	z-value	p-value
antonym-related	1.0564	0.101	10.449	<0.001
antonym-unrelated	0.0801	0.116	0.688	0.7703
related-unrelated	-0.9763	0.101	-9.690	<0.001

Marginal trends were computed using marginal means. The Tukey method was used to adjust p-values for three comparisons.

the absence of semantic relatedness in the violation condition: unrelated non-antonyms are easily categorized as a mismatch because there is no overlap in semantic features with the expected antonyms. Hence, the SAT profile seems to be mainly determined by the ease of categorizing the perceived input as an antonym, rather than by the processing of (predictable or semantically related) linguistic properties *per se*.

Hence, one can conclude that, in the antonym paradigm, processing semantic relatedness—as revealed by reductions in N400 amplitude—does not influence behavioral signatures in a similar vein, i.e., it does not lead to faster or more accurate performance. Rather, semantic relatedness is an intervening factor for categorization, as we have suggested based on its negative effect on P300 amplitude (see “Experiment 1: Antonym Processing and ERPs” above). From this perspective, the SAT data seem more in line with the ERP data than standard RT measures.

Yet, with separate analyses we can still not directly relate the two data sets to each other. Therefore, we conducted a joint analysis of the SAT and EEG data to investigate whether behavioral performance was driven by N400-related processes, P300-related processes or both.

Modeling SAT Dynamics as a Function of ERP Data

In addition to the direct modeling of the SAT response as a function of condition, we can also model the SAT response

TABLE 6 | Generalized linear mixed effect model for the SAT data (Experiment 2) based on event-related potential (ERP) responses.

AIC	BIC	logLik	deviance	df.resid
7,293	7,367	-3,637	7,273	11,484
Scaled residuals:				
Min	1Q	Median	3Q	Max
-15.05	0.09	0.2	0.37	2.95
Random effects:				
Groups	Term	Std.Dev.	Corr	
item	(Intercept)	0.65		
subj	(Intercept)	2.81		
	logRT	0.45	-0.996	
Number of obs: 11,494, groups: item, 20; subj, 15.				
Fixed effects:				
	Estimate	Std. Error	z-value	Pr (> z)
(Intercept)	-11	0.83	-13	1.5e-38 ***
logRT	1.8	0.13	14	6.2e-43 ***
n400.fitted	12	1.7	7.2	7.5e-13 ***
p300.fitted	-28	2.8	-9.9	2.8e-23 ***
logRT:n400.fitted	-2	0.23	-8.7	4.1e-18 ***
logRT:p300.fitted	4.6	0.38	12	2.1e-33 ***

Dependent variable is the response accuracy, correspondingly the model family is binomial with a logit link. RT is the total reaction time, i.e., the response time latency plus the reaction time to that tone. The EEG predictors are the average fitted response extracted from the respective models for each component. Model fit by maximum likelihood using the Laplace approximation and the bobyqa optimizer.

as a function of the mean ERP from the EEG experiment. For this model, fitted values by condition were extracted from mixed-effects models for the P300 and N400 and then aggregated to yield a single value for each component in each condition. These values are then used instead of the categorical predictor in an otherwise identical mixed-effect model for the SAT response. The difference in item sets (the EEG item set was larger) and participants, as well as the aggregation step, ensure that these values are not merely fitting within experiment item or participant variation, but rather capturing population-level dynamics.

The resulting model (Table 6) is identical in fit to the model based on the categorical condition codes (see Figure 5 and the AIC and logLik values in Tables 4, 6). At first this may seem surprising, but this model has an identical number of parameters and differs in practice only in its design matrix that no longer codes condition directly but rather the electrophysiological “encoding” of (the response to) the condition. This decomposes the different processes present in each condition—much in the same way that independent components in ICA present the same data as the original channel-wise EEG yet reveal insights about latent structure.

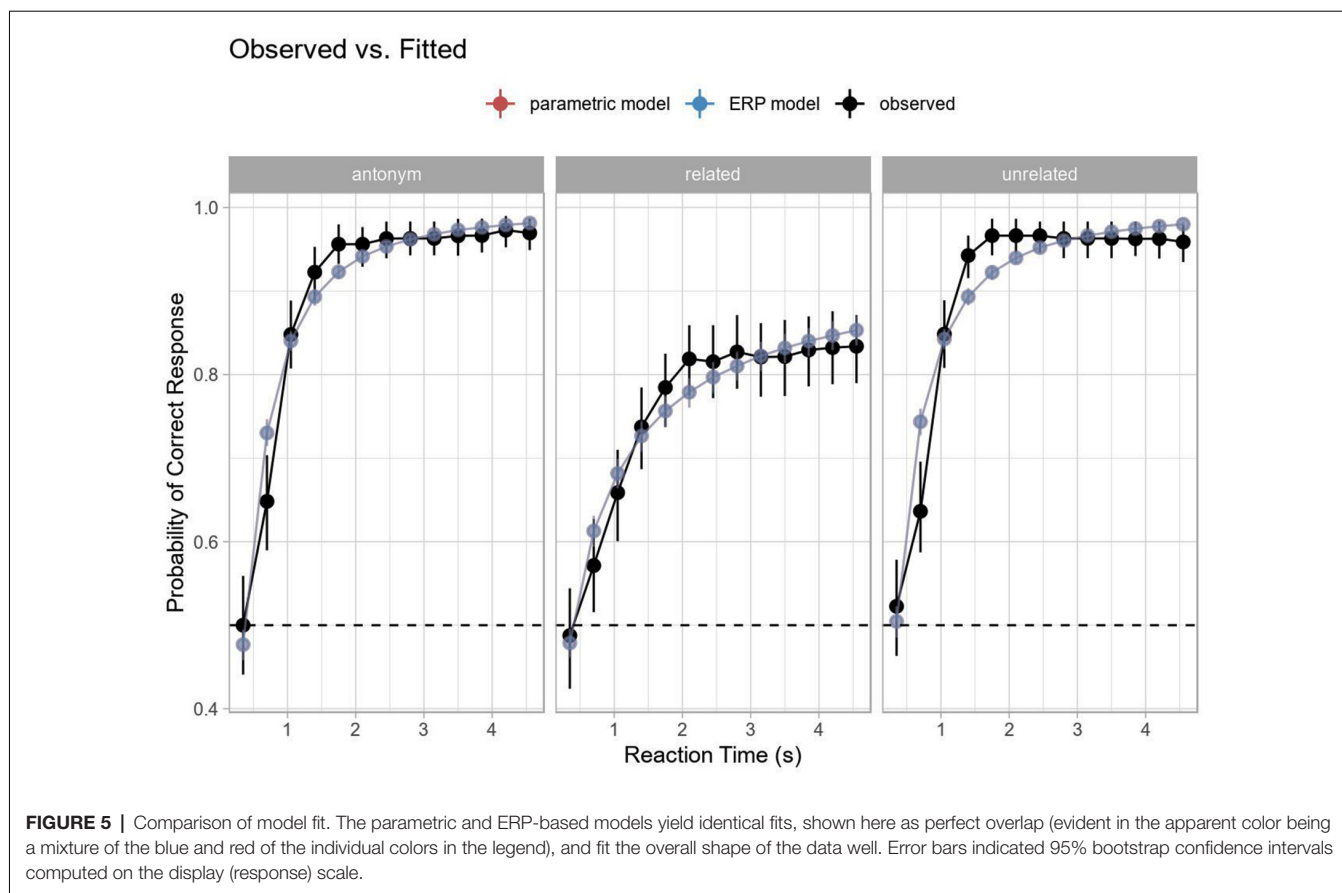
The partial effect plot in Figure 6 shows this most clearly. The curves for each component were obtained by removing the effect for the respective other component (by setting the corresponding predictor to zero using the remef package, Hohenstein and Kliegl, 2015). In the antonym condition, the P300 dominates and this reflects the dominant categorization process for a full prediction match. In the unrelated condition, the N400 dominates and reflects processing the complete prediction mismatch. In the related condition, the N400 is also the dominant effect, but less so, reflecting a mixture of matching (i.e., semantically related) and mismatching features. The partial effects for each individual component, but especially for the N400, make a further prediction for the related condition: both the predicted rate of increase towards terminal accuracy and the terminal

accuracy would have been lower than in the unrelated condition. In other words, the largest processing difficulties arise from stimuli that neither completely fulfill predictions nor are clear errors, even though such stimuli do not necessarily elicit the largest ERP components. Thus, this too is in line with the hypothesis sketched above that semantic relatedness interferes with antonym categorization.

Moreover, the main effect for the N400 response in the model reflects an increased probability of correct responses with a decreased N400 amplitude; the accompanying interaction effect with log RT shows that this effect decreases with longer response latencies (see the asymptotic behavior of the N400 curve in Figure 6). This may suggest that the processes underlying the N400 become more decoupled from the categorization process over time, which fits with our assumption that stimulus processing (as reflected in the N400) and categorization states (as reflected in the P300) are connected, yet distinct processes.

Meanwhile, the P300 shows the opposite effect: the main effect of P300 amplitude reflects an initially lower probability of correct response, while its interaction with log RT shows that P300 amplitude is associated with a higher probability of correct response as a function of time. This is compatible with previous research suggesting a decoupling of P300 peak latency and response accuracy at shorter response latencies (see Kutas et al., 1977) and with recent proposals that P300 activity, in general, may reflect the ongoing accumulation of evidence for subsequent decision-making (Twomey et al., 2015).

Overall, this shows that in the antonym design as implemented here, ERP responses to antonyms are indexing categorization dynamics with little influence from N400 activity, while the reverse holds for the two mismatch conditions. Our results also suggest that N400 and P300 responses show reversed influences on accuracy depending on response latency. With increasing response time, reduced N400 amplitudes predict correct responses to a lesser degree, whereas P300 is a worse predictor for response accuracy at shorter latencies. This suggests



that ERP-behavior links inferred from standard RT measures are likely to show variation depending on whether the behaviorally indexed decision point falls in early or late bins on the overall continuum of response times in a given experiment.

ANALYSIS OF COMPONENT OVERLAP IN THE EEG DATA

Throughout the article, and specifically in our modeling of the SAT response as a function of the average amplitude of the P300 and N400 components, we have assumed that these two components are largely independent, or at least two sides of the coin. Furthermore, while our chosen time windows reduce component overlap, they do not eliminate it. To better understand the relationship between the two components, we take a two-pronged approach, considering both the P300 amplitude as a covariate in predicting the N400 amplitude and a multivariate Bayesian model, which allows for modeling both components simultaneously in a single model. The analysis source code is available on the OSF (see “Data Availability Statement” below).

Using the P300 Amplitude as a Covariate in the N400 Model

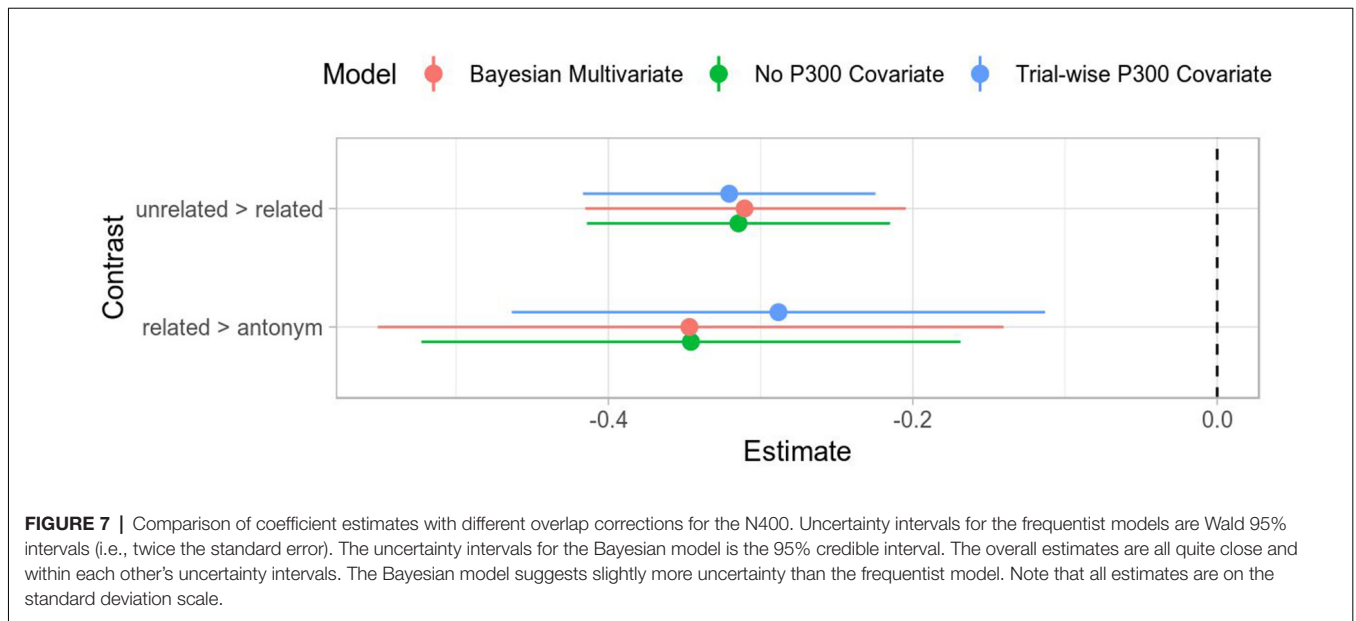
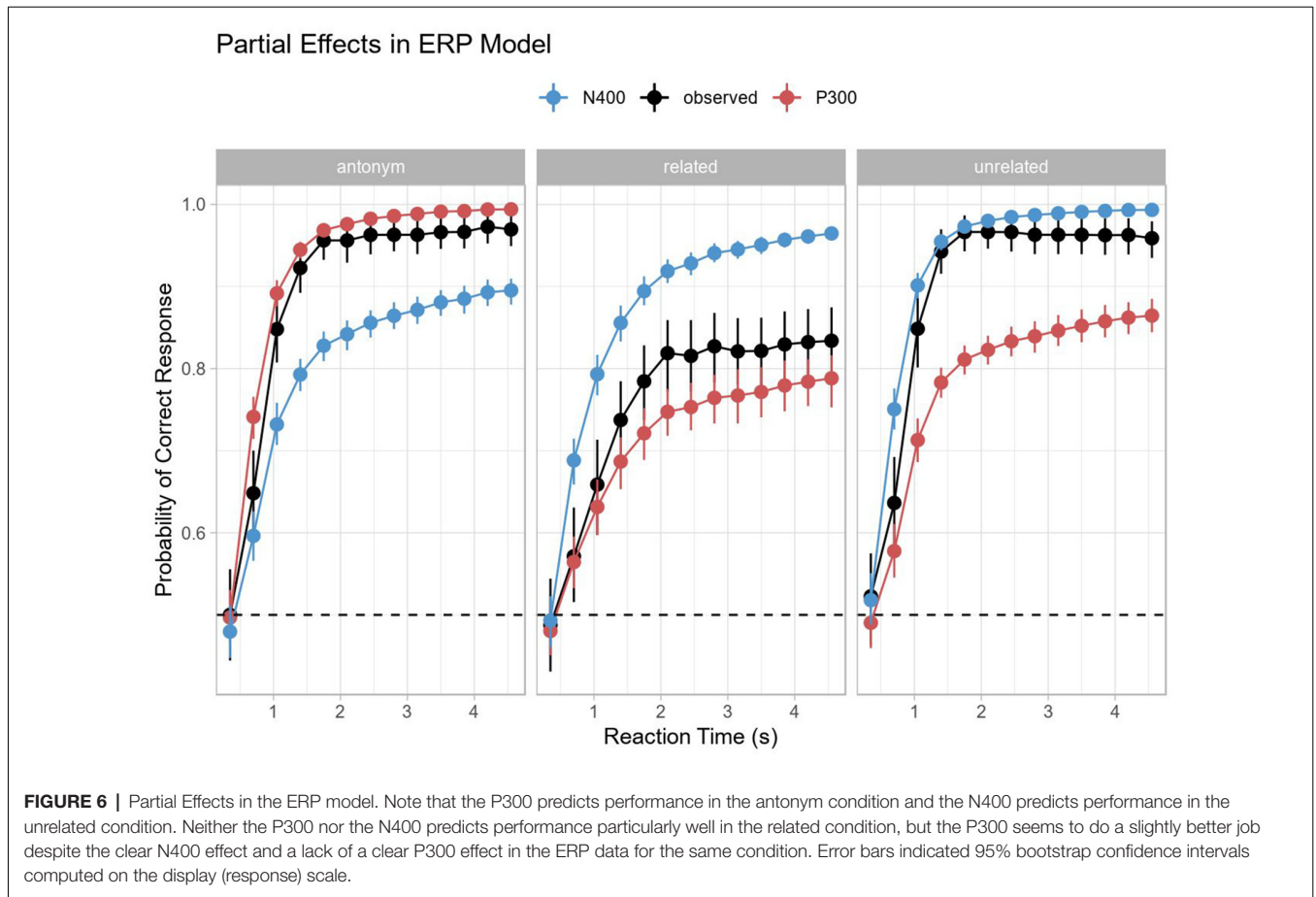
The simplest way to address component overlap is by including the scaled trial-wise P300 amplitude as a predictor for

the N400 amplitude as a main effect, which significantly improved model fit. Subsequent extension of this model by including all interaction terms did not significantly improve fit and so we prefer the simpler, more parsimonious model. Interestingly, neither the overall pattern of effects nor their numerical estimates changed much (see **Figure 7**), indicating that the P300 amplitude is an additive effect or offset for the N400 amplitude. The lack of an interaction effect and similar estimates for the other contrasts suggest that there is some component overlap in the N400 time window, but that the observed effects are independent of the effects in the P300 time window.

Although it may seem backward in time, we can also repeat this covariate analysis for the P300. This would accommodate for a rising N400 already occurring and overlapping with the P300 in the P300 time window. We again find that the overall model fit is better but that the effect is additive and does not greatly change our contrasts of interest (see **Figure 8**).

Bayesian Multivariate Model

Including the trial-wise P300 amplitude in the model for the N400 shows that our N400 effects are not strongly influenced by the preceding P300 (even if the total amplitude in the N400 time window is). However, we can go beyond treating the P300 as an offset for the N400 and jointly model both



effects using multivariate Bayesian mixed-effects models with brms (v2.7.0) and Stan (v2.18.2; Bürkner, 2017, 2018; Stan Development Team, 2018). In simple terms, these models can

be thought of distinct, simultaneous models that nonetheless inform each other, much in the same way that different groups in a mixed-effect model inform each other *via* partial pooling.

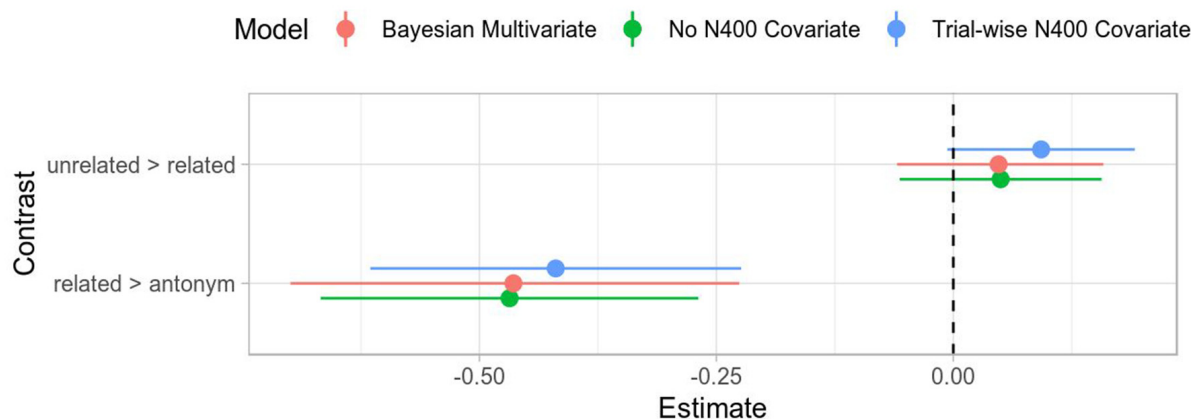


FIGURE 8 | Comparison of coefficient estimates with different overlap corrections for the P300. Uncertainty intervals for the frequentist models are Wald 95% intervals (i.e., twice the standard error). The uncertainty intervals for the Bayesian model is the 95% credible interval. The overall estimates are all quite close and within each other's uncertainty intervals. The Bayesian model suggests slightly more uncertainty than the frequentist model. Note that all estimates are on the standard deviation scale.

This information sharing across submodels furthermore allows for examining covariance between shared predictors for multiple dependent variables and more directly reflects the intertwined nature of the data. In other words, it allows examining how effects are related across different dependent variables. This is similar to structural equation modeling; indeed, it is possible to compute many structural equation models this way.

Given that the frequentist results suggest that including the P300 amplitude as a covariate does not greatly impact our effect estimates in the N400 time window, we omit it from the multivariate model for computational efficiency. As in the EEG analysis above, we use the mean voltage in the baseline window as well as condition as fixed-effect predictors. Our random effects are identical to the analysis above (see “Analysis and Results” section), but with an additional correlation level for the by-item and by-subject effects across dependent measures. Our dependent measures are simultaneously the P300 and N400 responses. All variables are coded and transformed as above.

No priors were set on the random effects beyond the default, which yields point estimates for the random effects comparable to lme4. For the fixed effects, a normal prior with mean of 0 and standard deviation of 2 was used. This is a lightly regularizing prior, equal to the assumption that most effects are small (68% are less than two standard deviations in size) and nearly all are not large (95% are less than four standard deviations in size). This is analogous to weakly-penalized ridge (L2-regularized) regression in frequentist estimation.

The model was fit using Markov Chain Monte Carlo and the No-U-Turn-Sampler (Homan and Gelman, 2014), a self-tuning variant of Hamiltonian Monte Carlo. For all parameters, the Gelman-Rubin statistic R_{hat} was equal to 1.0 and the number of effective samples exceeded 4,000; for the condition contrasts, the number of effective samples exceeded 7,500. A full model summary can be found in the **Supplementary Materials**.

The correlation of the by-subject random effects across response variables was not distinguishable from zero (the credible interval crossed zero for all pairwise correlations). This suggests that between-subject variation in the P300 response is not noticeably correlated with the between-subject variation in the N400 response. The correlation of residuals between the different response variables was small but non zero (credible interval: 0.05–0.13). This suggests that there is shared residual variation in both components that is not captured by our predictors.

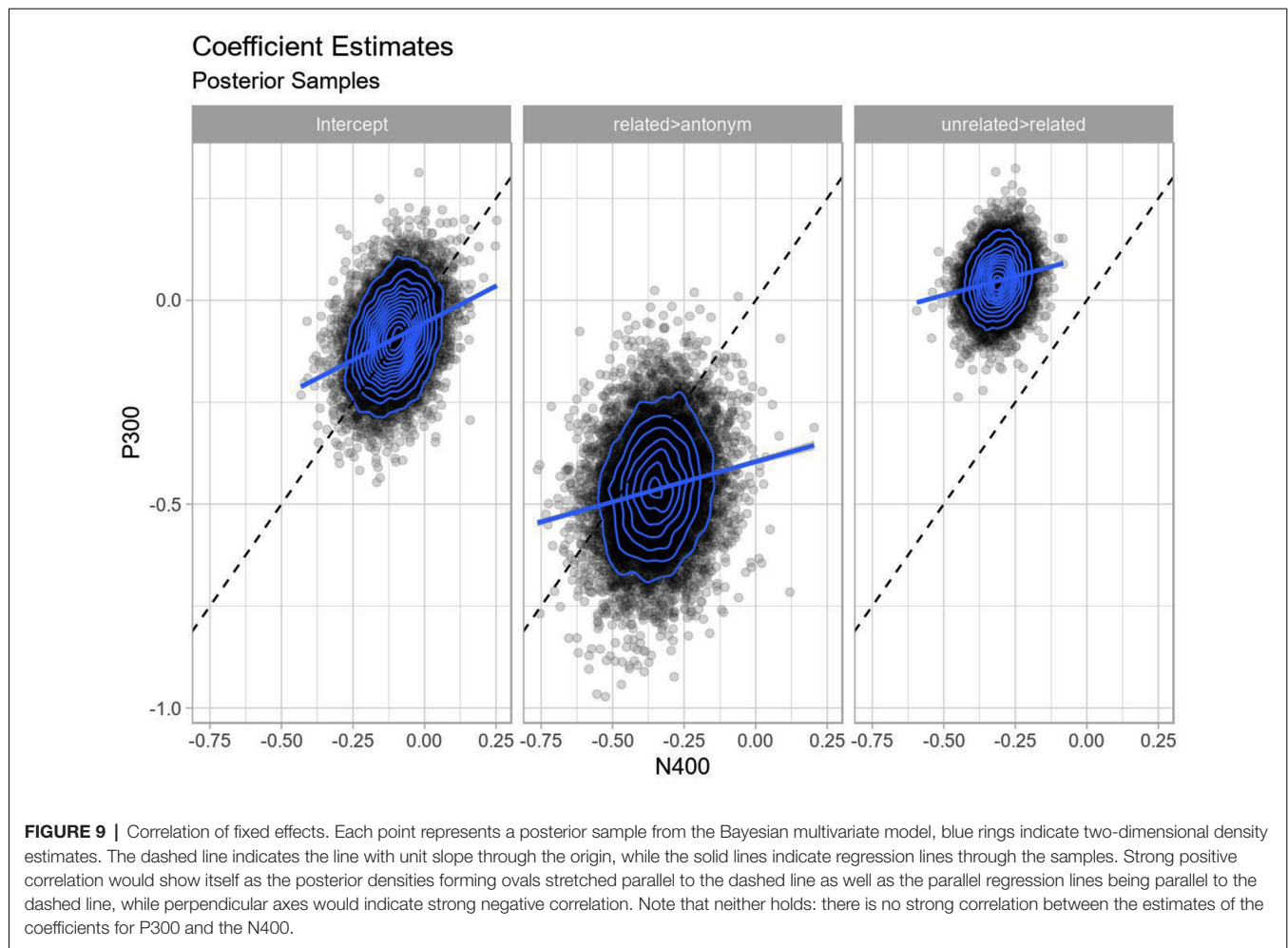
The correlation for the fixed effects between components was always positive, but generally small (Pearson correlation of 0.12–0.31; see also **Figure 9**). This corresponds to some component overlap—a positive deflection from a P300 will shift the basis for the negative deflection for N400 in the positive direction, much like the additive offset behavior in the frequentist model—but does not correspond to completely dependent components, where we would expect stronger collinearity.

Finally, the overall estimates for all effects are similar to the univariate analyses above, although with a larger uncertainty for the related > antonym contrast, reflecting a somewhat larger uncertainty in component-wise amplitude differences between strongly P300-evoking and the strongly N400-evoking conditions (see **Figures 7, 8**).

Taken together, the frequentist covariate models for each component and the Bayesian multivariate model provide converging evidence for the observed effects for each component being independent of each other and not profoundly distorted by temporal overlap.

GENERAL DISCUSSION

The present article revisited a long-standing issue in the EEG literature on language processing, namely the relationship



between multidimensional, time-sensitive electrophysiological data and unidimensional, time-insensitive behavioral data. We hypothesized that previous investigations on this issue faced two methodological challenges: the inherent ambiguity in offline RT measures, conflating response speed, accuracy and different kinds of online processes, and the temporal (and topographical) component overlap of endogenous ERPs such as the N400 and P300. In dealing with the first challenge, we proposed that using time-sensitive behavioral measures such as the SAT paradigm may moderate interpretative ambiguity of RT measures resulting from only observing a single snapshot of completed processing. As for the second challenge, we proposed that cross-method mixed-effects models may be a feasible solution. We examined these issues with the antonym paradigm that has yielded conflicting ERP and behavioral results as well as a strong overlap of N400 and P300 responses to target words.

In terms of the interpretive ambiguity of standard RT measures, we found that time-sensitive behavioral measures can provide more insightful data. Specifically, the SAT data showed that unexpected non-antonym targets that were related to the correct antonym exhibited lowest terminal accuracy and slowest increase in accuracy. This pattern of results is compatible with

the view that semantic relatedness of an unexpected sentence completion hinders categorization by sharing semantic features with the expected antonym or, equivalently, overlapping in along a different categorization axis (e.g., for the word pair *black-yellow* this would be the feature of being a color term). In line with the interference assumption proposed for semantic relatedness, we did not find a significant difference between antonyms and unrelated non-antonyms in their terminal accuracy nor the trajectory towards it. This clearly contrasts with the results reported previously where the unrelated condition was processed significantly different than the antonym condition (Bentin, 1987; Roehm et al., 2007; Federmeier et al., 2010). Given that RT measures using a single button press constitute just one data point on a SAT curve, one may speculate that the contrast observed previously fell within an RT range in which the differences between the two conditions were most pronounced, while failing to capture dynamic development between earlier and later bins with indistinguishable asymptotes. One way to test this possibility would be to use varying latencies between target word and decision prompts in future ERP experiments on antonym processing, i.e., merging a single-response SAT design (e.g., McElree et al., 2003) with EEG collection. If

done carefully, this would also allow for the separation of stimulus- and response-locked components, an aspect that, due to experimental setup, we could not address in our treatment of the P300 (and N400) responses. One open question for further research is whether response-locked components may be a better predictor of SAT responses, thereby also revealing whether it makes a difference to categorize a prediction match or mismatch.

Our modeling of SAT responses as a function of EEG activity lends further support to the hypothesis that standard RT measures may be measuring different spots along the SAT curve (also across conditions), which is not under experimental control. We found interactive effects for response time and ERP activity as predictors of response accuracy. Importantly, while reductions in N400 amplitude were a better predictor for response accuracy at shorter latencies, the reversed pattern held for the P300. Surely, any inference as to which ERP component influences response accuracy obtained with standard RT measures will depend on where on the hypothesized SAT curve that RT data point will be positioned. As argued above, this can be accounted for by systematically sampling the latencies between RT measures and target processing or, as already proposed by others, by modeling accuracies as a function of response time (e.g., Davidson and Martin, 2013). Finally, our modeling approach also attests to the feasibility that ERP responses in one sample predict behavioral SAT responses in another, and may therefore be particularly suitable for experimental designs, where the specifics of the single-response SAT procedure appear impossible to be combined with EEG recordings for practical reasons (e.g., due to the higher number of experimental trials needed to compute a robust signal, resulting in an excessive number of experimental sessions). In general, the modeling technique proposed here also applies to combining EEG with further behavioral methods, such as eye-tracking or skin conductance, that may necessitate partly different experimental designs than EEG setups to guarantee internal validity.

Regarding ERP component overlap in time, we hypothesized that the N400 and P300 responses during linguistic categorization show related, yet distinguishable processes. Specifically, we conjectured that the N400 would be more sensitive to processing stimulus properties relevant for categorization (including linguistic fit, see Bornkessel-Schlesewsky and Schlesewsky, 2019), while the P300 indexes the dynamics of the categorization process itself (O'Connell et al., 2012; Twomey et al., 2015). Component overlap is a notorious problem in interpreting ERP patterns, as it makes it extremely difficult to determine whether amplitude modulations in a given component under study are the result of offsets introduced by an adjacent component (additive component overlap), reflective of modulations within a given component or a mixture of the two (multiplicative component overlap). One way to address this problem in the case of overlapping N400 and P300 responses is to deploy the attested sensitivity of the P300 to task variation and associated attention orientation. That is, naturalistic tasks (e.g., reading or listening for comprehension) or tasks that direct participants' attention away from stimulus properties used for

linguistic categorization help reduce P300 overlap (e.g., Roehm et al., 2007; Haupt et al., 2008). Yet, task variation may not always be an option for various reasons, the most obvious one being that categorization itself is of interest. The present article takes the extreme version of the opposite end of task variation: using a behavioral measure to help disentangle components. Our choice of behavioral task and stimulus paradigm elicits a strong categorization response (P300) independent of a response to the congruency and fit of the stimulus (N400). This results in ERP effects that we can separate statistically and which provide a useful basis for decomposing and understanding processing time-courses as exhibited behaviorally in the SAT paradigm. In other words, understanding the perception-action loop can be better understood when we manipulate both perception and action.

In summary, the current experiments and analyses strongly suggest that combining EEG with time-sensitive behavioral measures from SAT designs enriches our understanding of both ERPs elicited by language input and the resulting behavioral performance in categorization tasks. The SAT data suggest that, in the antonym paradigm, N400 priming effects due to semantic relatedness do not affect behavioral performance, unless they impact negatively on categorization, whereas categorization processes clearly dominate response behavior. As a consequence, the current SAT data can be integrated more readily with explanations of the possible cognitive functions of the N400 (stimulus-related processes) and the P300 (categorization dynamics).

In our modeling approach, we have restricted ourselves to two components and their temporal overlap to demonstrate the feasibility of this type of cross-method analyses. There are several possibilities of how our modeling approach can be extended in future research. First, recall that our data sets are based on a stimulus paradigm that yields near-perfect cloze probability for the predicted target word. An obvious application is to test the current approach with experimental designs inducing a broader range of cloze probability values to measure predictability. Second, our modeling approach can be applied to other types of component mixtures as well. This includes not only topographic overlap of distinct ERPs, but also temporal overlap of the N400 and the ensuing late positivity. Throughout the present article, for instance, we have argued that the positivity in response to non-antonyms indexes eventual categorization for prediction mismatches, hence is also a P300 with a latency shift. Follow-up studies could test to what extent late positivities in other experimental designs overlap with or are independent of the N400, thereby also further testing assumptions on the nature of the late positivity (see Leckey and Federmeier, 2019).

CONCLUSION

We presented here a novel application of modern statistical approaches to better understand the complex interaction between behavior and electrophysiology and more generally between offline and online measures. We demonstrated a general technique for combining data from multiple methods, resulting in a novel decomposition of competing neural processes

underlying behavior. Subsequently, we used a combination of techniques to disentangle two classically entwined ERP effects, the P300 and N400, with potential applications to other component mixtures. To see the dynamics of processing in its full depth, we must examine distinct measures together, much in the same way that depth perception arises from combining two distinct perspectives. Only in the combination of perception and action do we see the full loop and thus, by closing the perception-action loop, we learn more about both perception and action.

DATA AVAILABILITY

The datasets generated for this study (preprocessed EEG data and SAT data) along with analysis source code on the Open Science Framework (OSF) will be made available upon publication of this manuscript. For review purposes, these may be viewed at https://osf.io/75r6t/?view_only=b1c45caff34a48558b580e8a7202cfe7.

ETHICS STATEMENT

Experiment 1 was conducted in the XLinc Lab at the University of Cologne. The protocol for ERP experiments conducted in the lab is approved by the Ethics Committee of the German Society of Linguistics (DGfS; #2016-09-160914). Experiment 2 was not accompanied by an ethics vote but was conducted in line with national and institutional guidelines. Specifically, behavioral non-invasive experiments with healthy young adults (between 18–65 years) do not require one as long as they pose no risk or physical/emotional burden to participants and as long as participants are debriefed after participation as specified by the rules of the German Research Foundation (DFG; https://www.dfg.de/foerderung/faq/geistes_sozialwissenschaften/), archived

REFERENCES

- Alday, P. M. (2017). How much baseline correction do we need in ERP research? Extended GLM model can replace baseline correction while lifting its limits. *Psychophysiology* 00:e13451. doi: 10.1111/psyp.13451
- Alday, P. M., Schlesewsky, M., and Bornkessel-Schlesewsky, I. (2017). Electrophysiology reveals the neural dynamics of naturalistic auditory language processing: event-related potentials reflect continuous model updates. *eNeuro* 4:ENEURO.0311-16.2017. doi: 10.1523/eneuro.0311-16.2017
- Bakker, I., Takashima, A., van Hell, J. G., Janzen, G., and McQueen, J. M. (2015). Tracking lexical consolidation with ERPs: lexical and semantic-priming effects on N400 and LPC responses to newly-learned words. *Neuropsychologia* 79, 33–41. doi: 10.1016/j.neuropsychologia.2015.10.020
- Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *J. Mem. Lang.* 68. doi: 10.1016/j.jml.2012.11.001
- Bates, D., Kliegl, R., Vasishth, S., and Baayen, H. (2015a). Parsimonious mixed models. *arXiv [Preprint]*. <https://arxiv.org/abs/1506.04967>. Accessed May 26, 2018.
- Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015b). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01
- Bentin, S. (1987). Event-related potentials, semantic processes, and expectancy factors in word recognition. *Brain Lang.* 31, 308–327. doi: 10.1016/0093-934x(87)90077-0

page on 27 Feb 2019 available at https://web.archive.org/web/20190227214057/https://www.dfg.de/foerderung/faq/geistes_sozialwissenschaften/).

AUTHOR CONTRIBUTIONS

FK designed the experimental protocol and ran the experiments. PA analyzed the data. FK and PA wrote the manuscript.

FUNDING

FK acknowledges funding from the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) as part of the CRC 1252 “Prominence in Language” in the project B07 “Agentivity as a key to prominence: Experimental approaches to argument alternations in German” at the University of Cologne.

ACKNOWLEDGMENTS

We thank R. Muralikrishnan for his invaluable help in setting up the presentation script for Experiment 2. We also thank Tim Graf and Brita Rietdorf for help in data acquisition for Experiment 1 and Brita Rietdorf, Miriam Burk and Elisabeth Beckermann for help in data acquisition for Experiment 2. Experiment 2 was conducted while FK was at the University of Mainz and the University of Marburg, Germany.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnhum.2019.00285/full#supplementary-material>

- Bornkessel, I., McElree, B., Schlesewsky, M., and Friederici, A. D. (2004). Multi-dimensional contributions to garden path strength: dissociating phrase structure from case marking. *J. Mem. Lang.* 51, 495–522. doi: 10.1016/j.jml.2004.06.011
- Bornkessel-Schlesewsky, I., Kretzschmar, F., Tune, S., Wang, L., Genc, S., Philipp, M., et al. (2011). Think globally: cross-linguistic variation in electrophysiological activity during sentence comprehension. *Brain Lang.* 117, 133–152. doi: 10.1016/j.bandl.2010.09.010
- Bornkessel-Schlesewsky, I., Philipp, M., Alday, P. M., Kretzschmar, F., Grew, T., Gumpert, M., et al. (2015). Age-related changes in predictive capacity versus internal model adaptability: electrophysiological evidence that individual differences outweigh effects of age. *Front. Aging Neurosci.* 7:217. doi: 10.3389/fnagi.2015.00217
- Bornkessel-Schlesewsky, I., and Schlesewsky, M. (2019). Towards a neurobiologically plausible model of language-related, negative event-related potentials. *Front. Psychol.* 10:298. doi: 10.3389/fpsyg.2019.00298
- Bott, L., Bailey, T. M., and Grodner, D. (2012). Distinguishing speed from accuracy in scalar implicatures. *J. Mem. Lang.* 66, 123–142. doi: 10.1016/j.jml.2011.09.005
- Bourguignon, N., Drury, J., Valois, D., and Steinhauer, K. (2012). Decomposing animacy reversals between agents and experiencers: an ERP study. *Brain Lang.* 122, 179–189. doi: 10.1016/j.bandl.2012.05.001
- Burkhardt, P. (2006). Inferential bridging relations reveal distinct neural mechanisms: evidence from event-related brain potentials. *Brain Lang.* 98, 159–168. doi: 10.1016/j.bandl.2006.04.005

- Bürkner, P.-C. (2017). brms: an R package for bayesian multilevel models using stan. *J. Stat. Softw.* 80, 1–28. doi: 10.18637/jss.v080.i01
- Bürkner, P. (2018). Advanced bayesian multilevel modeling with the R package brms. *R J.* 10, 395–411. doi: 10.32614/rj-2018-017
- Chwilla, D. J., Kolk, H. H., and Mulder, G. (2000). Mediated priming in the lexical decision task: evidence from event-related potentials and reaction time. *J. Mem. Lang.* 42, 314–341. doi: 10.1006/jmla.1999.2680
- Collins, A. M., and Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychol. Rev.* 82, 407–428. doi: 10.1037/0033-295x.82.6.407
- Davidson, D., and Martin, A. E. (2013). Modeling accuracy as a function of response time with the generalized linear mixed effects model. *Acta Psychol.* 144, 83–96. doi: 10.1016/j.actpsy.2013.04.016
- Debrulle, J. B., Rodier, M., Prévost, M., Lionnet, C., and Molavi, S. (2013). Effects of a small dose of olanzapine on healthy subjects according to their schizotypy: an ERP study using a semantic categorization and an oddball task. *Eur. Neuropsychopharmacol.* 23, 339–350. doi: 10.1016/j.euroneuro.2012.06.005
- Degno, F., Loberg, O., Zang, C., Zhang, M., Donnelly, N., and Liversedge, S. P. (2019). Parafoveal previews and lexical frequency in natural reading: evidence from eye movements and fixation-related potentials. *J. Exp. Psychol. Gen.* 148, 453–474. doi: 10.1037/xge0000494
- Dimigen, O., Sommer, W., Hohlfield, A., Jacobs, A. M., and Kliegl, R. (2011). Coregistration of eye movements and EEG in natural reading: analyses and review. *J. Exp. Psychol. Gen.* 140, 552–572. doi: 10.1037/a0023885
- 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
- Federmeier, K. D., and Kutas, M. (1999). A rose by any other name: long-term memory structure and sentence processing. *J. Mem. Lang.* 41, 469–495. doi: 10.1006/jmla.1999.2660
- Federmeier, K. D., Kutas, M., and Schul, R. (2010). Age-related and individual differences in the use of prediction during language comprehension. *Brain Lang.* 115, 149–161. doi: 10.1016/j.bandl.2010.07.006
- Gold, J. I., and Shadlen, M. N. (2007). The neural basis of decision making. *Annu. Rev. Neurosci.* 30, 535–574. doi: 10.1146/annurev.neuro.29.051605.113038
- Graf, T., Philipp, M., Xu, X., Kretzschmar, F., and Primus, B. (2017). The interaction between telicity and agentivity: experimental evidence from intransitive verbs in German and Chinese. *Lingua* 200, 84–106. doi: 10.1016/j.lingua.2017.08.006
- Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., et al. (2013). MEG and EEG data analysis with MNE-Python. *Front. Neurosci.* 7:267. doi: 10.3389/fnins.2013.00267
- Haupt, F. S., Schlesewsky, M., Roehm, D., Friederici, A. D., and Bornkessel-Schlesewsky, I. (2008). The status of subject-object reanalyses in the language comprehension architecture. *J. Mem. Lang.* 59, 54–96. doi: 10.1016/j.jml.2008.02.003
- Hohenstein, S., and Kliegl, R. (2015). remef: Remove Partial Effects. R package version 1.0.6.9000. Available online at: <https://github.com/hohenstein/remef/>. Accessed August 15, 2019.
- 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 Kounios, J. (1990). “Semantic memory: an ERP and reaction time analysis.” in *Psychophysiological Brain Research*, eds C. H. M. Brunia, A. W. K. Gaillard and A. Kok (Tilburg: Tilburg University Press), 285–288.
- Holcomb, P. J., and Neville, H. (1990). Auditory and visual semantic priming in lexical decision: a comparison using event-related brain potentials. *Lang. Cogn. Processes* 5, 281–312. doi: 10.1080/01690969008407065
- Homan, M. D., and Gelman, A. (2014). The No-U-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res.* 15, 1593–1623.
- Johns, C. L., Matsuki, K., and Van Dyke, J. A. (2015). Poor readers’ retrieval mechanism: efficient access is not dependent on reading skill. *Front. Psychol.* 6:1552. doi: 10.3389/fpsyg.2015.01552
- Johnson, R. (1986). A triarchic model of P300 amplitude. *Psychophysiology* 23, 367–384. doi: 10.1111/j.1469-8986.1986.tb00649.x
- Johnson, B. W., and Hamm, J. P. (2000). High-density mapping in an N400 paradigm: evidence for bilateral temporal lobe generators. *Clin. Neurophysiol.* 111, 532–545. doi: 10.1016/s1388-2457(99)00270-9
- Kelly, S. P., and O’Connell, R. G. (2015). The neural processes underlying perceptual decision making in humans: recent progress and future directions. *J. Physiol. Paris* 109, 27–37. doi: 10.1016/j.jphysparis.2014.08.003
- Kiefer, M. (2001). Perceptual and semantic sources of category-specific effects: event-related potentials during picture and word categorization. *Mem. Cognit.* 29, 100–116. doi: 10.3758/bf03195745
- Kounios, J., and Holcomb, P. J. (1992). Structure and process in semantic memory: Evidence from event-related brain potentials and reaction times. *J. Exp. Psychol.* 121, 459–479. doi: 10.1037/0096-3445.121.4.459
- Kretzschmar, F. (2010). *The Electrophysiological Reality of Parafoveal Processing: On the Validity of Language-Related ERPs in Natural Reading*. Marburg: University of Marburg. Unpublished doctoral dissertation. Available online at: <https://archiv.ub.uni-marburg.de/diss/z2011/0602/pdf/dfk.pdf>. Accessed August 15, 2019.
- Kretzschmar, F., Bornkessel-Schlesewsky, I., and Schlesewsky, M. (2009). Parafoveal versus foveal N400s dissociate spreading activation from contextual fit. *Neuroreport* 20, 1613–1618. doi: 10.1097/wnr.0b013e328332c4f4
- Kretzschmar, F., Schlesewsky, M., and Staub, A. (2015). Dissociating word frequency and predictability effects in reading: evidence from coregistration of eye movements and EEG. *J. Exp. Psychol. Learn. Mem. Cogn.* 41, 1648–1662. doi: 10.1037/xlm0000128
- 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 between the lines: Event-related brain potentials during natural sentence processing. *Brain Lang.* 11, 354–373. doi: 10.1016/0093-934X(80)90133-9
- Kutas, M., and Iragui, V. (1998). The N400 in a semantic categorization task across 6 decades. *Electroencephalogr. Clin. Neurophysiol.* 108, 456–471. doi: 10.1016/s0168-5597(98)00023-9
- Kutas, M., McCarthy, G., and Donchin, E. (1977). Augmenting mental chronometry: the P300 as a measure of stimulus evaluation time. *Science* 197, 792–795. doi: 10.1126/science.887923
- Kutas, M., Neville, H. J., and Holcomb, P. J. (1987). A preliminary comparison of the N400 response to semantic anomalies during reading, listening and signing. *Electroencephalogr. Clin. Neurophysiol. Suppl.* 39, 325–330.
- Laszlo, S., and Federmeier, K. D. (2009). A beautiful day in the neighborhood: an event-related potential study of lexical relationships and prediction in context. *J. Mem. Lang.* 61, 326–338. doi: 10.1016/j.jml.2009.06.004
- Lau, E., Phillips, C., and Poeppel, D. (2008). A cortical network for semantics: (de)constructing the N400. *Nat. Rev. Neurosci.* 9, 920–933. doi: 10.1038/nrn2532
- Leckey, M., and Federmeier, K. D. (2019). The P3b and P600(s): positive contributions to language comprehension. *Psychophysiology* doi: 10.1111/psyp.13351 [Epub ahead of print].
- Löbner, S. (2013). *Understanding Semantics*. 2nd Edn. London and New York: Routledge.
- Lotze, N., Tune, S., Schlesewsky, M., and Bornkessel-Schlesewsky, I. (2011). Meaningful physical changes mediate lexical-semantic integration: top-down and form-based bottom-up information sources interact in the N400. *Neuropsychologia* 49, 3573–3582. doi: 10.1016/j.neuropsychologia.2011.09.009
- Martin, A. E., and McElree, B. (2009). Memory operations that support language comprehension: evidence from verb-phrase ellipsis. *J. Exp. Psychol. Learn. Mem. Cogn.* 35, 1231–1239. doi: 10.1037/a0016271
- Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., and Bates, D. (2017). Balancing Type I error and power in linear mixed models. *J. Mem. Lang.* 94, 305–315. doi: 10.1016/j.jml.2017.01.001
- McElree, B. (2006). Accessing recent events. *Psychol. Learn. Motiv.* 46, 155–200. doi: 10.1016/s0079-7421(06)46005-9

- McElree, B., Foraker, S., and Dyer, L. (2003). Memory structures that subserve sentence comprehension. *J. Mem. Lang.* 48, 67–91. doi: 10.1016/s0749-596x(02)00515-6
- Meade, G., and Coch, D. (2017). Word-pair priming with biased homonyms: N400 and LPC effects. *J. Neurolinguistics* 41, 24–37. doi: 10.1016/j.jneuroling.2016.09.002
- Neely, J. H. (1991). “Semantic priming effects in visual word recognition: a selective review of current findings and theories,” in *Basic Progresses in Reading-Visual Word Recognition*, eds D. Beser and G. W. Humphreys (Hillsdale, NJ: Erlbaum), 264–333.
- Nieuwenhuis, S., Aston-Jones, G., and Cohen, J. D. (2005). Decision making, the P3 and the locus coeruleus-norepinephrine system. *Psychol. Bull.* 131, 510–532. doi: 10.1037/0033-2909.131.4.510
- Nieuwenhuis, S., De Geus, E. J., and Aston-Jones, G. (2011). The anatomical and functional relationship between the P3 and autonomic components of the orienting response. *Psychophysiology* 48, 162–175. doi: 10.1111/j.1469-8986.2010.01057.x
- Nieuwland, M. S., Barr, D. J., Bartolozzi, F., Busch-Moreno, S., Darley, E., Donaldson, D. I., et al. (2019). Dissociable effects of prediction and integration during language comprehension: evidence from a large-scale study using brain potentials. *bioRxiv* [Preprint]. doi: 10.1101/267815
- O’Connell, R. G., Dockree, P. M., and Kelly, S. P. (2012). A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. *Nat. Neurosci.* 15, 1729–1735. doi: 10.1038/nn.3248
- 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
- Pfefferbaum, A., Ford, J., Johnson, R. Jr., Wenegrat, B., and Kopell, B. S. (1983). Manipulation of P3 latency: speed vs. accuracy instructions. *Electroencephalogr. Clin. Neurophysiol.* 55, 188–197. doi: 10.1016/0013-4694(83)90187-6
- Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clin. Neurophysiol.* 118, 2128–2148. doi: 10.1016/j.clinph.2007.04.019
- Rabovsky, M., and McRae, K. (2014). Simulating the N400 ERP component as semantic network error: insights from a feature-based connectionist attractor model of word meaning. *Cognition* 132, 68–89. doi: 10.1016/j.cognition.2014.03.010
- Roehm, D. (2004). *Waves and Words: Oscillatory Activity and Language Processing*. Marburg, Germany: University of Marburg. Unpublished doctoral dissertation. Available online at: <https://archiv.ub.uni-marburg.de/diss/z2005/0110/pdf/ddr.pdf>. Accessed August 15, 2019.
- Roehm, D., Bornkessel-Schlesewsky, I., Rösler, F., and Schlewsky, M. (2007). To predict or not to predict: influences of task and strategy on the processing of semantic relations. *J. Cogn. Neurosci.* 19, 1259–1274. doi: 10.1162/jocn.2007.19.8.1259
- Rolke, B., Heil, M., Streb, J., and Hennighausen, E. (2001). Missed prime words within the attentional blink evoke an N400 semantic priming effect. *Psychophysiology* 38, 165–174. doi: 10.1111/1469-8986.3820165
- Sassenhagen, J., and Bornkessel-Schlesewsky, I. (2015). The P600 as a correlate of ventral attention network reorientation. *Cortex* 66, A3–A20. doi: 10.1016/j.cortex.2014.12.019
- Stan Development Team. (2018). *RStan: The R Interface to Stan. R Package Version 2.18.2*. Available online at: <http://mc-stan.org/>. Accessed August 15, 2019.
- Tanner, D., Morgan-Short, K., and Luck, S. J. (2015). How inappropriate high-pass filters can produce artifactual effects and incorrect conclusions in ERP studies of language and cognition. *Psychophysiology* 52, 997–1009. doi: 10.1111/psyp.12437
- 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
- Twomey, D. M., Murphy, P. R., Kelly, S. P., and O’Connell, R. G. (2015). The classic P300 encodes a build-to-threshold decision variable. *Eur. J. Neurosci.* 42, 1636–1643. doi: 10.1111/ejn.12936
- van Berkum, J. J. V., Hagoort, P., and Brown, C. M. (1999). Semantic integration in sentences and discourse: evidence from the N400. *J. Cogn. Neurosci.* 11, 657–671. doi: 10.1162/089892999563724
- Van Petten, C., and Luka, B. (2012). Prediction during language comprehension: benefits, costs and ERP components. *Int. J. Psychophysiol.* 83, 176–190. doi: 10.1016/j.ijpsycho.2011.09.015
- Verleger, R., Jaśkowski, P., and Wascher, E. (2005). Evidence for an integrative role of P3b in linking reaction to perception. *J. Psychophysiol.* 19, 165–181. doi: 10.1027/0269-8803.19.3.165
- Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta Psychol.* 41, 67–85. doi: 10.1016/0001-6918(77)90012-9

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Alday and Kretzschmar. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Identifying the Speech Production Stages in Early and Late Adulthood by Using Electroencephalography

Jakolien den Hollander¹, Roel Jonkers², Peter Mariën^{3,4†} and Roelien Bastiaanse^{2,5*}

¹ International Doctorate in Experimental Approaches to Language and Brain (IDEALAB, Universities of Groningen, Potsdam, Newcastle, Trento and Macquarie University), Sydney, NSW, Australia, ² Center for Language and Cognition Groningen (CLCG), University of Groningen, Groningen, Netherlands, ³ Clinical and Experimental Neurolinguistics (CLIEN), Vrije Universiteit Brussel, Brussels, Belgium, ⁴ Department of Neurology and Memory Clinic, ZNA Middelheim General Hospital, Antwerp, Belgium, ⁵ Center for Language and Brain, National Research University Higher School of Economics, Moscow, Russia

OPEN ACCESS

Edited by:

Yury Y. Shtyrov,
Aarhus University, Denmark

Reviewed by:

Vasil Kolev,
Institute of Neurobiology (BAS),
Bulgaria
Evangelos Paraskevopoulos,
Aristotle University of Thessaloniki,
Greece

*Correspondence:

Roelien Bastiaanse
y.r.m.bastiaanse@rug.nl

[†] Peter Mariën passed away on November 01, 2017. He took the initiative for this project. Without him the current study could not have been performed.

Specialty section:

This article was submitted to
Speech and Language,
a section of the journal
Frontiers in Human Neuroscience

Received: 30 January 2019

Accepted: 12 August 2019

Published: 10 September 2019

Citation:

den Hollander J, Jonkers R,
Mariën P and Bastiaanse R (2019)
Identifying the Speech Production
Stages in Early and Late Adulthood by
Using Electroencephalography.
Front. Hum. Neurosci. 13:298.
doi: 10.3389/fnhum.2019.00298

Structural changes in the brain take place throughout one's life. Changes related to cognitive decline may delay the stages of the speech production process in the aging brain. For example, semantic memory decline and poor inhibition may delay the retrieval of a concept from the mental lexicon. Electroencephalography (EEG) is a valuable method for identifying the timing of speech production stages. So far, studies using EEG mainly focused on a particular speech production stage in a particular group of subjects. Differences between subject groups and between methodologies have complicated identifying time windows of the speech production stages. For the current study, the speech production stages lemma retrieval, lexeme retrieval, phonological encoding, and phonetic encoding were tracked using a 64-channel EEG in 20 younger adults and 20 older adults. Picture-naming tasks were used to identify lemma retrieval, using semantic interference through previously named pictures from the same semantic category, and lexeme retrieval, using words with varying age of acquisition. Non-word reading was used to target phonological encoding (using non-words with a variable number of phonemes) and phonetic encoding (using non-words that differed in spoken syllable frequency). Stimulus-locked and response-locked cluster-based permutation analyses were used to identify the timing of these stages in the full time course of speech production from stimulus presentation until 100 ms before response onset in both subject groups. It was found that the timing of each speech production stage could be identified. Even though older adults showed longer response times for every task, only the timing of the lexeme retrieval stage was later for the older adults compared to the younger adults, while no such delay was found for the timing of the other stages. The results of a second cluster-based permutation analysis indicated that clusters that were observed in the timing of the stages for one group were absent in the other subject group, which was mainly the case in stimulus-locked time windows. A z-score mapping analysis was used to compare the scalp distributions related to the stages between the older and younger adults. No differences between both groups were observed with respect to scalp distributions, suggesting that the same groups of neurons are involved in the four stages, regardless of the adults' age, even though the timing of the individual stages is different in both groups.

Keywords: speech production, aging, electroencephalography, word retrieval, articulation

INTRODUCTION

Effects of Aging on the Brain

Structural changes in the brain, such as a reduction in cortical thickness (Freeman et al., 2008; Zheng et al., 2018), a decrease in the number of cortical folds (Zheng et al., 2018), and a reduction in gray (Freeman et al., 2008) and white matter (Marner et al., 2003) take place throughout one's lifetime. Also, the connectivity within the cingulo-opercular network [CON; including dorsal anterior cingulate, medial superior frontal cortex, anterior insula, frontal operculum, and anterior prefrontal cortex (Dosenbach et al., 2007)] and the frontoparietal control network [FPCN; including the lateral prefrontal cortex, anterior cingulate cortex, and inferior parietal lobule (Vincent et al., 2008)] reduces with aging (Geerligns et al., 2015). These networks modulate higher cognitive functions involved in language processing, such as working memory and reading. While the global efficiency of the three networks is the same in older and younger adults, the local efficiency and the modularity decrease with aging. This decrease may delay the speech production process; however, the efficiency of the visual network, which is used when watching pictures, is maintained. Therefore, no delay in the processing of information has been observed in the visual network with aging.

Age-related changes in the brain are also reflected in the oscillations of the brain, which can be measured using electroencephalography (EEG). The amplitude of components (peaks that are related to a particular process in the brain) in the processed signal, observed when many neurons fire together, is reduced in older individuals (Wlotko et al., 2010). There are two reasons why this reduction may occur: (1) neurons that fire together are geometrically less aligned and do no longer fire synchronously and (2) the latency of the component is more variable. Also, delays in the latency of the N400 component have been observed in older individuals. According to the global slowing hypothesis (Brinley, 1965), older adults are slower in every process, which should be reflected in the EEG. Slower processing speed may, thus, be observed in older adults when carrying out a cognitive task, because they cannot focus on speed when they are focusing on responding as accurately as possible, known as the "speed-accuracy tradeoff" (Ratcliff et al., 2007). Not being able to focus on both speed and accuracy is possibly related to a decrease in the strength of the tract between the presupplementary motor area and the striatum in older adults (Forstmann et al., 2011).

Effects of Aging on the Speech Production Process

Between 25 and 100% of the structural and functional changes in the brain are related to cognitive decline (Fjell and Walhovd, 2011). Cognitive decline caused by aging may have an effect on the speech production process. For example, older adults are less accurate in picture naming than younger adults (Connor et al., 2004). Decline in object naming is accompanied by a reduction in white and gray matter in the left temporal lobe (Cardenas et al., 2011). The temporal lobe has been associated with semantic memory, in which concepts are stored. When a

concept activates a lemma (the word meaning) in the lexicon, semantically related lemmas get coactivated. The correct lemma is retrieved from the mental lexicon when lemmas that are semantically related to the target are sufficiently inhibited. Both semantic memory and inhibition decline with aging (Harada et al., 2013).

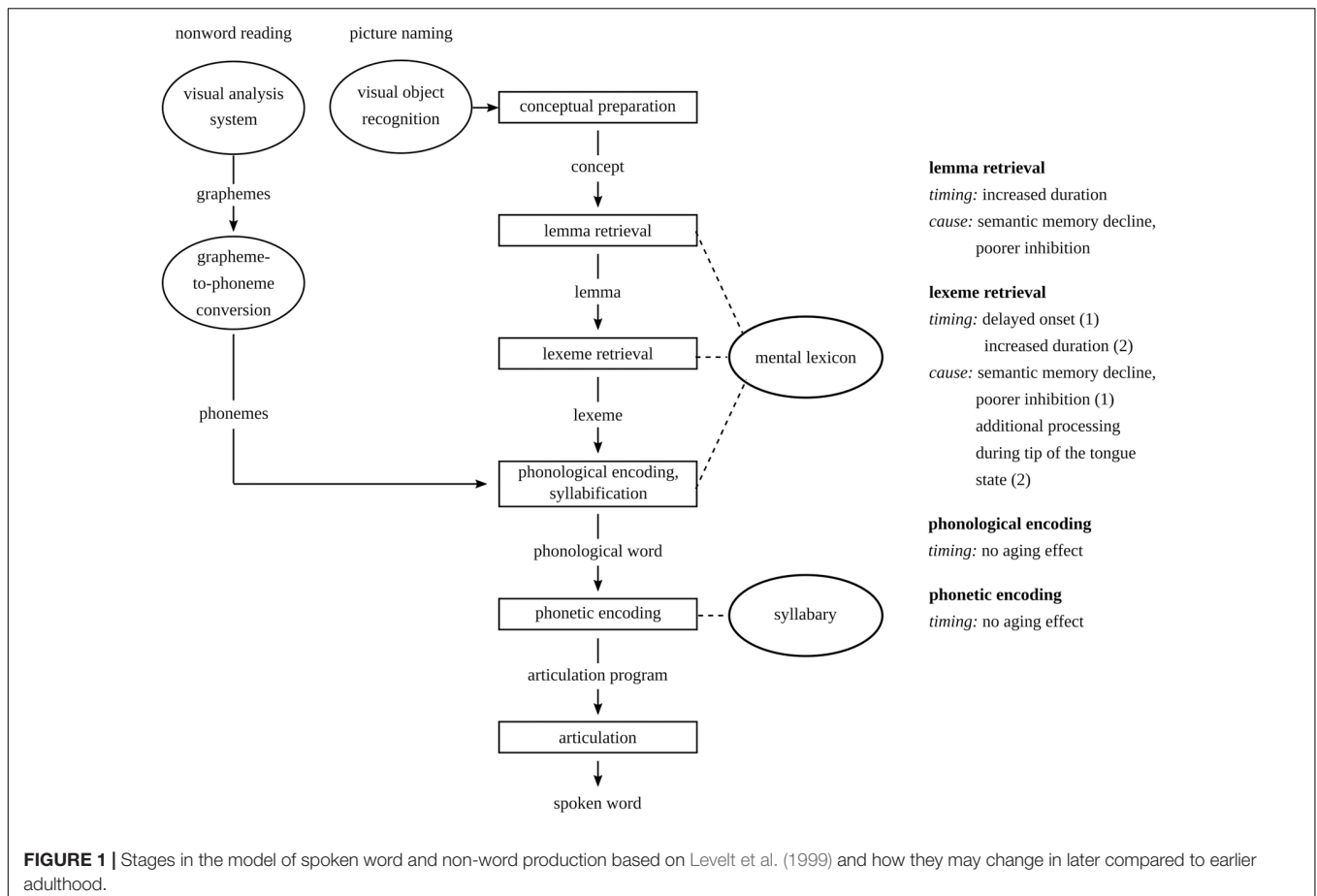
After the lemma retrieval stage, the lexical word form, the lexeme, is retrieved. When there is insufficient information available about the lexeme, the phonological form of the word cannot be retrieved. The speaker experiences a temporal failure to produce a word even though the word is well known to him. This so-called tip-of-the-tongue phenomenon is observed more frequently in older adults, particularly in those with atrophy in the left insula (Shafto et al., 2007).

In the next stage of object naming, phonological encoding, the phonemes corresponding to the lexeme are retrieved and ordered and the phonological rules are applied. No aging effects have been reported for phonological encoding. Finally, the string of phonemes is phonetically encoded into an articulation plan. This plan specifies how the muscles of the mouth and throat will interact during the articulation of the word. Older individuals have a longer response duration for the production of both sequential and alternating syllable strings, which is associated with reduced cortical thickness in the right dorsal anterior insula and in the left superior temporal sulcus and gyrus (Tremblay and Deschamps, 2016).

In sum, delayed lemma retrieval can be observed in older individuals (Cardenas et al., 2011) due to reduced semantic memory and poorer inhibition abilities (Harada et al., 2013). A delay at the lemma level may delay the onset of lexeme retrieval. Lexeme retrieval may be delayed due to tip-of-the-tongue states (Shafto et al., 2007). In this study, lemma and lexeme retrieval are studied in picture-naming tasks, while phonological and phonetic encoding are studied in non-word production tasks. Since lemma and lexeme retrieval do not play a role in non-word production tasks, delays in these stages cannot delay the onset of phonological and phonetic encoding. Aging is not expected to have an effect on these two stages, because no aging effects on phonological encoding have been reported. Also, the task used to study phonetic encoding is different from the task used by Tremblay and Deschamps (2016). An overview of the stages in spoken word and non-word production that may change in later adulthood is provided in **Figure 1**.

Current Study

The hypothesis that the lemma and lexeme retrieval stages are delayed in older compared to younger individuals, whereas phonological and phonetic encoding are similar in both groups, can be tested using EEG. Since each speech production stage has its own timing (Indefrey, 2011), it is possible to identify the individual stages using tasks in which more processing is required at the particular stage. Lemma retrieval requires more effort when the number of previously retrieved lemmas from neighboring nodes increases. This effect is referred to as the "cumulative semantic interference effect" (Howard et al., 2006). Two EEG studies have used this effect to target the stage of lemma retrieval, which has been identified from 150 to 225 ms



(Maess et al., 2002) and from 200 to 380 ms after stimulus presentation (Costa et al., 2009).

Lexeme retrieval requires more effort when the age of acquisition (AoA) of words increases (Laganaro and Perret, 2011; Laganaro et al., 2012; Valente et al., 2014). This stage has been identified in a time window from 120 to 350 ms after stimulus presentation and around 280 and 150 ms before response onset (Laganaro and Perret, 2011), from 380 to 400 ms after stimulus presentation and up to 200 ms before response onset (Laganaro et al., 2012), and from 380 after stimulus presentation up to 100 ms before response onset (Valente et al., 2014).

Phonological encoding requires more effort when the number of phonemes increases. So far, word length effects have not been identified in EEG studies, meaning that the time frame of phonological encoding has not been identified yet using this manipulation (Valente et al., 2014; Hendrix et al., 2017). However, other tasks, such as comparing overt and covert production of nouns and verbs, have been used to track phonological encoding (Sahin et al., 2009). In the current study, non-word length is used, which may lead to different findings.

Syllable frequency is known to have an effect on phonetic encoding: when syllable frequency decreases, phonetic encoding requires more effort (Levelt and Wheeldon, 1994). In a task in which phonemes were inserted into non-words with varying frequencies in a non-word reading task, the syllable frequency

effect has been identified using EEG from 170 to 100 ms before response onset (Bürki et al., 2015). Our methodology is different because participants were asked to read the non-words, not to insert phonemes. It is, therefore, unclear what to expect.

Hence, for the current study, the cumulative semantic interference effect, the AoA effect, the effect of non-word length in phonemes, and the syllable frequency effect will be used to track the speech production stages in a group of younger adults and in a group of older adults. The time windows of the stages in both groups will be identified. If the time windows of the stages differ between the two groups, that does not mean that the processing mechanisms are different (Nieuwenhuis et al., 2011). Therefore, a direct comparison of both groups will be made in the time windows of the relevant stages that were identified in the younger adults and the older adults. Additionally, the scalp distributions of the stages will be compared between the two groups.

MATERIALS AND METHODS

Participants

For the group of young adults, 20 young adulthood native speakers of Dutch (5 males) participated. The mean age of the participants was 21.8 years (age range: 17–28 years). Participants

in the group of older adults were 20 late adulthood native speakers of Dutch (7 males). Their average age was 55.4 years (range: 40–65). The young adult participants are referred to as “younger adults,” and the late adulthood participants are referred to as “older adults.” The younger adults’ data will be the basis of this study, and their data will be compared to those of the older adults.

All participants were right handed, measured using the short version of the Edinburgh Handedness Inventory (Oldfield, 1971). They reported no problems in hearing, and their vision was normal or corrected to normal. Also, they reported no reading difficulties. All participants were financially compensated and gave informed consent. The study was approved by the *Ethics Committee of Humanities* of the University of Groningen.

Materials

Lemma Retrieval

The materials used in the lemma retrieval task were black-and-white drawings. The pictures originated from the *Auditief Taalbegripsprogramma (ATP; Bastiaanse, 2010)* and the *Verb and Action Test (VAT; see Bastiaanse et al., 2016)* for individuals with aphasia. The order in which the depicted nouns were presented was manipulated for the cumulative semantic interference effect. The pictures were grouped in sets of five semantically related neighbors (e.g., bed, couch, cradle, closet, and chair) that fit into a particular category (e.g., furniture, clothes, and insects). The five nouns within one category had the same number of syllables and the same stress pattern and were controlled for logarithmic lemma frequency in Dutch (Baayen et al., 1995). The depicted nouns were all mono- or disyllabic in Dutch.

For the selection of the final item list, a picture-naming task was carried out by four participants (one male) with a mean age of 22 years (age range: 21–23 years). Items that were named incorrectly by more than one participant were removed. The 125 selected items had an overall name agreement of 91.4%. The overall mean logarithmic lemma frequency was 1.28 (range: 0–2.91). The same set of pictures was used in two lists with reversed conditions to avoid an order of appearance effect. The lists were presented in three blocks of 30 items and one block of 35 items.

The pictures were presented on a computer screen, and participants were asked to name the pictures as quickly and accurately as possible. Before the picture was presented, a black fixation cross on a white background was shown for 500 ms. The function of the fixation cross was to draw attention and to announce that a picture was presented soon. The picture was shown for 5 s. Items within one category were not presented directly after another.

Lexeme Retrieval

The pictures for this test originated from the same sources as the materials on the first test and represented mono- and disyllabic nouns in Dutch. Items were controlled for AoA (Brysbaert et al., 2014) and lexeme frequency (Baayen et al., 1995).

Four participants (one male) with a mean age of 20.7 years (age range: 19–22) took part in a picture-naming task for pretesting the materials. These participants had not taken part in the lemma

retrieval task. Items that were named incorrectly by more than one participant were omitted.

The 140 selected items had an overall name agreement of 93.9%. AoA ranged from 4.01 years for the noun “book” to 9.41 years for the noun “anchor,” with a mean of 5.96 years. The mean logarithmic lexeme frequency was 1.02 (range: 0–2.44). The correlation between AoA and lexeme frequency in the items is significant [$r(138) = -0.28, p < 0.001$]. Therefore, in the analysis, only AoA has been taken into account. The items were organized in one list including four blocks of 35 items. The order of the items was randomized per block, so that every participant named the items in a different order.

The procedure of the lexeme retrieval task was the same as the procedure of the lemma retrieval task. Since there was some item overlap between the lemma and lexeme retrieval tasks, the two tasks were never administered consecutively. A non-word task was always administered in between.

Phonological and Phonetic Encoding

To identify the stages of phonological and phonetic encoding, a non-word reading task was used.¹ All non-words were disyllabic and composed of existing Dutch syllables. The combination of the two syllables resulted in a non-word, e.g., “kikkels” or “raalkro.” The non-words were controlled for spoken syllable frequency (Nederlandse Taalunie, 2004). Two lists of non-words were developed in written form for the reading task. The two lists contained the same syllables, but the syllables were combined differently; thus, the non-words were unique.

The non-words were pretested in a reading task by four participants who took part in pretesting the picture-naming tasks as well. Each list was pretested with two participants. The 140 selected items for list 1 had an accuracy rate of 100%; 8% of the non-words in list 2 were produced incorrectly. The syllables used in these items were combined into new non-words. These non-words were pretested again with two other participants. Their accuracy was 100%.

For each non-word, the average spoken syllable frequency was computed over its two syllables. For list 1, the mean frequency was 1,136 (range: 257–4,514) and 1,077 (range: 257–4,676) for list 2. Also, the number of phonemes in the non-words was controlled for, because the duration of phonological encoding may increase with the number of phonemes. For both lists, the number of phonemes in the non-words ranged from 3 to 8. The average number of phonemes was 5.33 for list 1 and 5.29 for list 2.

The non-words were presented in white letters on a black background. The font type Trebuchet MS Regular, size 64, was used. The stimulus was presented for 5 s and preceded by a fixation cross, which was presented for 500 ms. Participants read either list 1 or list 2. Each list was divided into four blocks of 35 items. The order in which the non-words was presented was randomized per block, so none of the participants read the non-words in the same order. The instruction was to read the non-words aloud as quickly and accurately as possible.

¹In fact, two non-word tasks were administered: reading and repetition. Since reading is more closely related to object naming (a visually presented stimulus evoking a spoken output), the data of the repetition task will be ignored.

General Procedure

During the experiments, participants were seated approximately 70 cm from the screen. E-Prime 2.0 (2012) was used to present the stimuli and to record the response times and the responses. A voice key was used to detect the response times. The responses were recorded using a microphone that was attached to a headset. Before the experiment started, participants practiced the task with five items for the picture-naming tasks and with eight items for the non-word reading task. Participants had the opportunity to take a short break between the four blocks of the experiments.

EEG Data Recording

Electroencephalography data were recorded with 128 (older adults) and 64 (younger adults) Ag/AgCl scalp electrodes (WaveGuard) cap using the EEGO and ASA-lab system (ANT Neuro Inc., Enschede, Netherlands). These systems are entirely compatible; EEGO is the latest version. For the older adults, only the 64 channels that were recorded in the younger group were analyzed. The full set of 128 electrodes was used in a different study. The electrode sites were distributed over the scalp according to the 10-10 system (Jasper, 1958) for the system with 64 electrodes and according to the 10-5 system for the system with 128 electrodes. Bipolar electrodes were used to record vertical ocular movements, such as eye blinks, for which the electrode sites were vertically aligned with the pupil and located above and below the left eye. Impedance of the skin was kept below 20 k Ω , which was checked before every experiment. Data were acquired with a sampling rate of 512 Hz, and reference was recorded from the mastoids.

Data Processing and Analysis

Behavioral Data

The audio recordings of the participants' responses were used to determine the speech onset time. The speech onset time in each audio file was manually determined using the waveform and the spectrogram in Praat (Boersma and Weenink, 2018). The speech onset times based on the audio files were used as response events in the response-locked EEG analysis. R was used for the statistical analysis of the behavioral and item data (R Core Team, 2017).

Trials to which participants responded incorrectly were excluded from the analysis (lemma retrieval: 7.8%; lexeme retrieval: 7.3%; phonological and phonetic encoding: 1.9%). Also, responses that included hesitations or self-corrections qualified as errors (lemma retrieval: 2.6%; lexeme retrieval: 2.6%; phonological and phonetic encoding: 0.8%). Items to which many participants responded extraordinarily fast or slow were excluded from the EEG analysis (lemma retrieval: 8%; lexeme retrieval: 18.6%; phonological and phonetic encoding: 12.1%). The average response time was computed over all accepted trials. Trials exceeding this average by 1.4 standard deviations were disregarded.

EEG Data

The EEG data were preprocessed using EEGLAB (Delorme and Makeig, 2004) as an extension to MATLAB (2015). After rereferencing to the average reference of the mastoids, the data were filtered with a 50-Hz notch filter to remove electricity noise

and bandpass filtered from 0.2 to 30 Hz. Then, the data were resampled to 128 Hz. Independent components analysis on all channels was used for artifact detection. Artifact components, such as eye blinks, were removed through visual inspection. Also, the effect of component removal on the data was visually inspected. The continuous data were segmented per trial from 200 ms until 2 s after stimulus onset. A baseline correction was applied over the data epochs, using the 200 ms before stimulus onset as a baseline. Then, the events of disregarded trials were removed. To study the time window from the stimulus onset until the response onset, both stimulus-locked analyses, in which the time window after stimulus onset is analyzed, and response-locked analyses, in which the backward time window before the response onset is analyzed, were carried out. For the stimulus-locked analysis, the data epochs were segmented from stimulus onset until one sampling point (8 ms) after the earliest response time. This one extra sampling point was removed before the analysis. The start of the response-locked analysis was determined by subtracting the stimulus-locked time window from the response onset. Depending on the task, accepted trials were coded into two or three conditions for the statistical analysis. The conditions are specified below per experiment. These data were exported from EEGLAB into the format used in FieldTrip (Oostenveld et al., 2011), which was used for the statistical analysis. Finally, the structure of the data files was prepared for a cluster-based permutation analysis (Maris and Oostenveld, 2007).

The aims of the analyses were to identify the time window of lemma retrieval with the cumulative semantic interference effect, the time window of lexeme retrieval with the AoA effect, the time window of phonological encoding with the non-word length in phonemes effect, and the time window of phonetic encoding with the syllable frequency effect. These time windows were identified in the group of older adults and in the group of younger adults using group-level cluster-based permutation analyses carried out over all participants per group. The cumulative semantic interference effect was computed as the difference between the first and the fifth presented item within a category. The difference between words with an AoA of around 5 years and words with an AoA of around 6 years, as well as the difference between words with an AoA of 5 years and words with an AoA of around 7 years were used to compute the AoA effect. The effect of non-word length in phonemes was computed as the difference between non-words consisting of four phonemes and non-words consisting of five phonemes, as well as the difference between non-words consisting of four phonemes and non-words consisting of six phonemes. The difference between non-words with a high syllable frequency of 1,000–1,500 and non-words with a moderate syllable frequency of 500–1,000, as well as the difference between non-words with a high syllable frequency of 1,000–1,500 and non-words with a low syllable frequency of 250–500 were used to compute the syllable frequency effect. In every analysis, the number of permutations computed was 5,000. The Monte Carlo method was used to compute significance probability, using a two-sided dependent samples *t*-test ($\alpha = 0.025$). In the first analysis of every experiment, the entire time window from stimulus onset until 100 ms before response onset was

tested. When an effect was revealed in this large time window, a smaller time window around the effect was tested once, so a more specific timing of the effect could be reported. Finally, the time windows of the stages in older and younger adults were compared. This method cannot show whether the two groups differ (Nieuwenhuis et al., 2011). Therefore, the EEGs of both groups have been compared in the time windows of the stages for every single condition using a cluster-based permutation analysis. Again, the Monte Carlo method was used to compute significance probability, but now a two-sided independent samples *t*-test ($\alpha = 0.025$) was used to compare the two subject groups.

Additionally, a *z*-score mapping analysis (Thatcher et al., 2002) was carried out to compare the scalp distributions of the older adults to those of the younger adults during the speech production stages. For each experiment, the data were analyzed in relevant time windows and conditions for which significant clusters were found in the cluster-based permutation analysis of the older and the younger adults. The length of these time windows varied between the participant groups, which would have caused a difference in the number of time points included in the analysis. To avoid this difference, the number of time points centered around the median of the longest time window used in the analysis was made equal to the number of time points in the shortest time window. For each time point, *z*-scores were computed per electrode. The mean computed over the younger adults' data was subtracted from each data point from the older adults' data individually. This subtraction was divided by the standard deviation computed over the younger adults' data. Mean *z*-scores were computed per condition. When the mean *z*-score deviated more than one standard deviation from zero, the difference between the age groups qualified as significant.

RESULTS

The mean, standard deviation, and range of the response time data from the three experiments are provided per participant group in **Table 1**. For all analyses on response time, only the correct responses were used.

Behavioral Results

Younger Adults

At all tasks, the younger adults performed at ceiling. The percentages of correct responses were 92.4% for lemma retrieval, 92.9% for lexeme retrieval, and 98% for the non-word reading task targeting phonological and phonetic encoding. On the lemma retrieval task, a cumulative semantic interference effect

was found on the response time [$F(1, 765) = 13.38, p < 0.001$]. Increased response times were found for pictures within a category that were presented at the fifth ordinal position compared to pictures that were presented at the first ordinal position. An AoA effect on the response time was identified on the lexeme retrieval task [$F(1, 2,205) = 104.01, p < 0.001$]. Response time increased as AoA advanced. Non-word length in number of phonemes is relevant at the level of phonological encoding and turned out to be a significant factor: response times increased when non-words consisted of more phonemes [$F(1, 2,096) = 5.71, p = 0.017$]. The frequency of the syllables was varied to tap into phonetic encoding. Response times were found to decrease when syllable frequency increased [$F(1, 2,320) = 6.35, p = 0.01$].

Older Adults

Like the younger adults, the older adults performed at ceiling on all tasks. The percentages of correct responses were 86.8% for lemma retrieval, 87.6% for lexeme retrieval, and 96.5% for the non-word reading tasks. A cumulative semantic interference effect was found on the lemma retrieval task [$F(1, 721) = 7.60, p = 0.006$]. Increased response times were found for pictures within a category that were presented at the fifth ordinal position compared to those presented at the first ordinal position. Also, increased response times were found for items with a later AoA on the task targeting lexeme retrieval [$F(1, 2,061) = 43.38, p < 0.001$]. In the non-word reading task, response times increased with the non-word length in number of phonemes, which was used as a marker for *phonological encoding* [$F(1, 1,943) = 5.60, p = 0.018$]. Furthermore, to target phonetic encoding, a decrease in syllable frequency of the non-words was found to increase response times [$F(1, 2,146) = 11.68, p < 0.001$].

Differences Between Younger and Older Adults

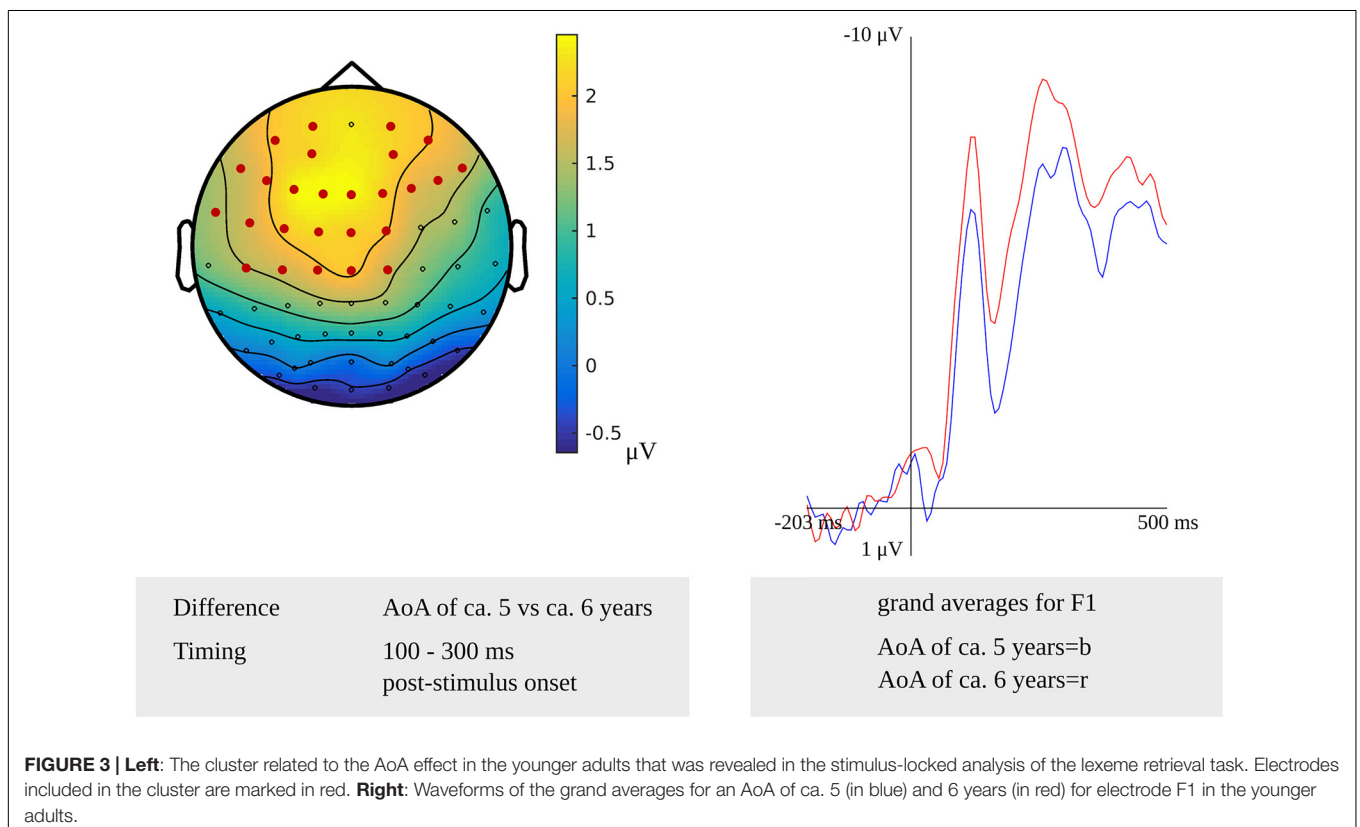
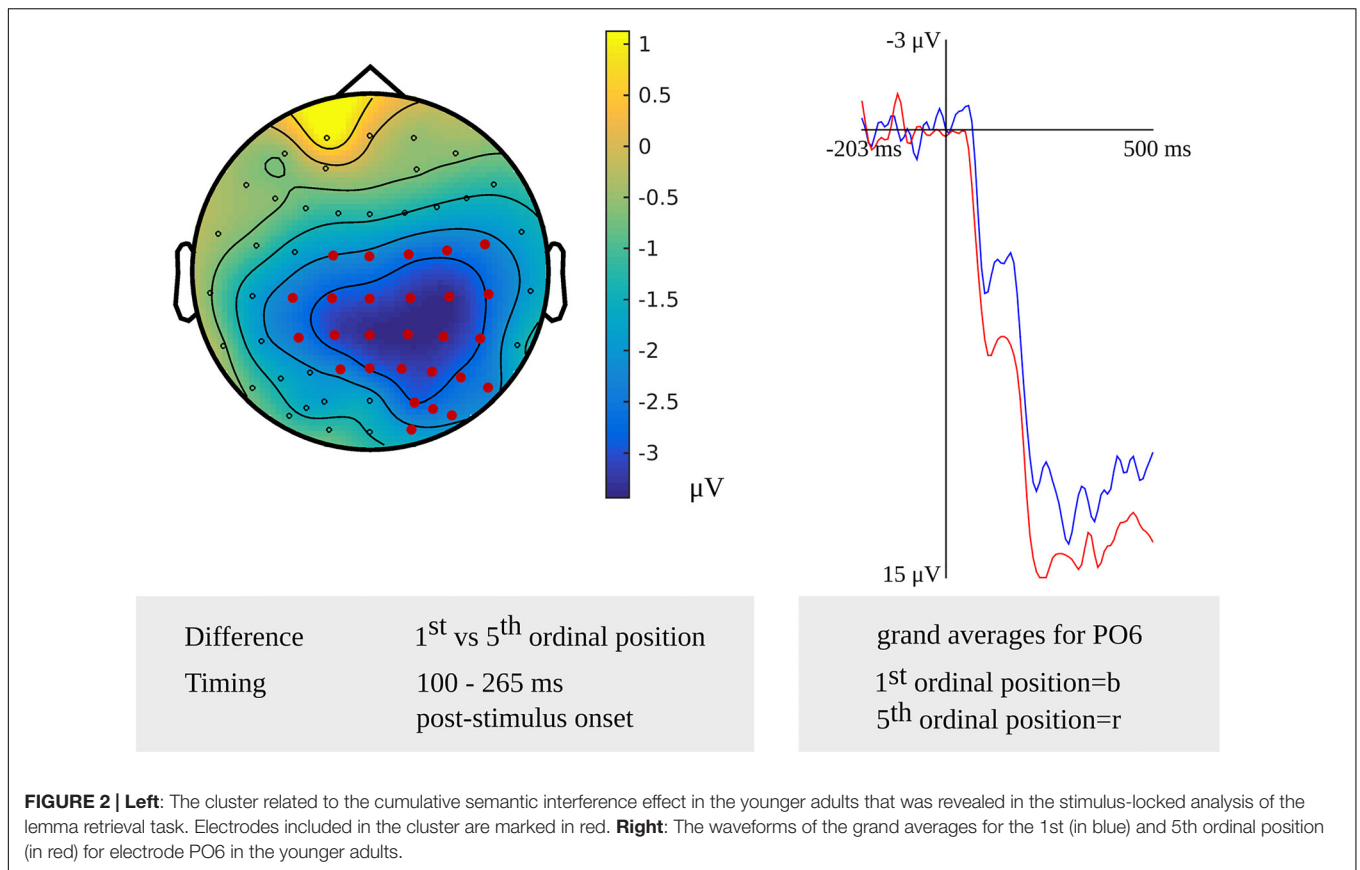
On *all* tasks, differences in response times between both age groups were found. The older adults responded slower than the younger adults on the lemma retrieval task [$F(1, 1,488) = 4.81, p = 0.028$], the lexeme retrieval task [$F(1, 4,268) = 7.14, p = 0.007$], and the non-word reading task targeting phonological and phonetic encoding [$F(1, 4,468) = 28.58, p < 0.001$]. Moreover, an interaction effect of AoA and participant age was found [$F(1, 4,268) = 4.51, p = 0.034$]. The group of older adults showed a smaller AoA effect [$F(1, 2,061) = 43.38, p < 0.001$] than the group of younger adults [$F(1, 2,205) = 104.01, p < 0.001$].

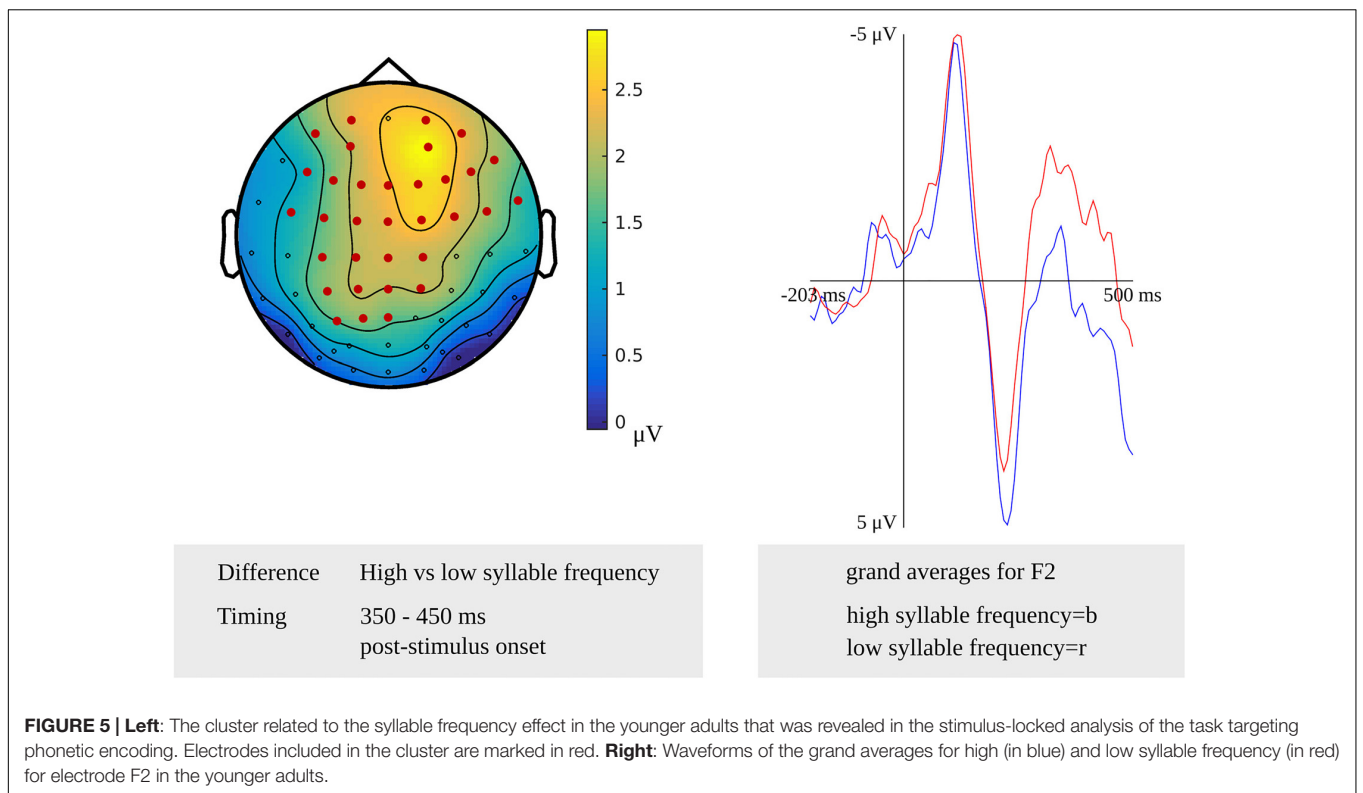
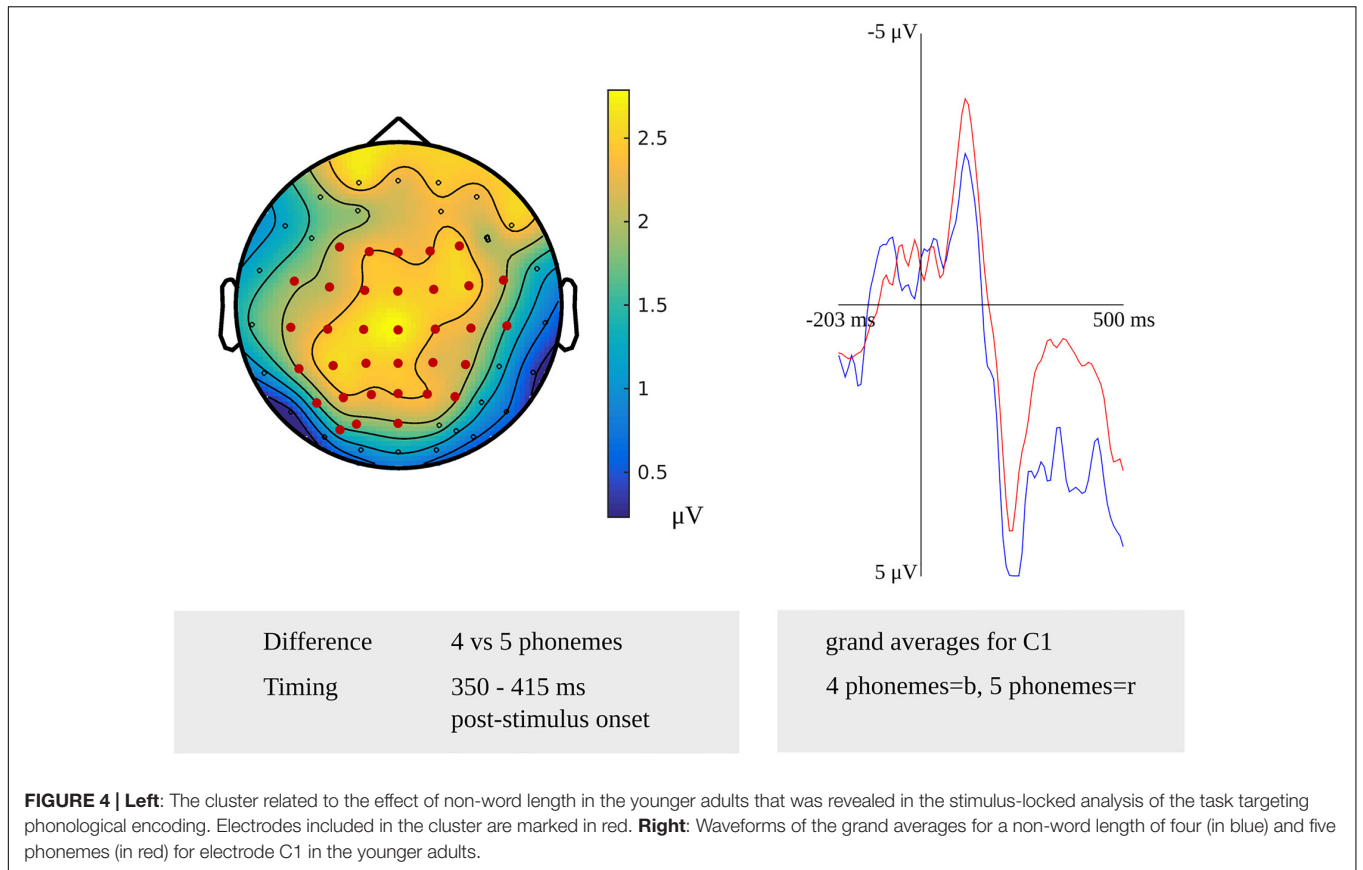
EEG Results

For the presentation of the EEG results, we will first present the results of the cluster-based permutation analysis for each task in

TABLE 1 | Response times of the younger and older adults.

Task	Mean (ms)		Standard deviation (ms)		Range (ms)	
	young	old	young	old	young	old
Lemma retrieval	932	944	216	213	602–1461	603–1460
Lexeme retrieval	938	946	199	201	626–1440	628–1439
Phonological and phonetic encoding in reading	690	699	116	119	502–966	504–965





the younger adults and then in the older adults to identify the time windows of the effects in these groups. Then, the differences between the two groups in these time windows computed with cluster-based permutation analyses will be presented along with the comparisons of the scalp distributions of both age groups. The EEG statistics are given in **Appendix 1A** (younger adults), **Appendix 1B** (older adults), and **Appendix 1C** (comparison of older and younger adults).

Younger Adults

In the younger adults, a difference between the first and fifth ordinal positions that was taken as evidence for the stage of lemma retrieval was revealed in the latency range from 100 to 265 ms ($p = 0.005$) after stimulus onset. The difference was most pronounced over right central and posterior sensors. In the response-locked analysis, an effect was found from 445 to 195 ms ($p = 0.004$) before response onset. The effect was most pronounced over central and posterior sensors bilaterally and over the right frontal electrodes. The scalp distribution of the stimulus-locked effect and the waveforms of the grand averages for the first and fifth ordinal position are shown in **Figure 2**.

Testing for an AoA effect targeting lexeme retrieval in the latency range from 100 to 300 ms after stimulus onset in the younger adults, the cluster-based permutation test revealed a difference between the items with an early AoA and items with a moderate AoA ($p = 0.002$). The difference was most pronounced on bilateral frontal and central sensors, as shown in **Figure 3**. **Figure 3** also shows the waveforms of the grand averages for the early and moderate AoA conditions. In the response-locked cluster-based permutation analysis, a difference between items with an early AoA and items with a late AoA was revealed from 475 to 330 ms before response onset. The response-locked AoA effect was most pronounced on bilateral frontal and bilateral central electrodes ($p < 0.001$).

A stimulus-locked length effect was revealed from 350 to 415 ms for the comparison of non-words consisting of four and five phonemes ($p = 0.0032$) targeting phonological encoding, which is shown in **Figure 4**. The waveforms of the grand averages for non-word length in four and five phonemes are provided in **Figure 4** as well. Also, a stimulus-locked length effect was revealed as a difference between non-words consisting of four and six phonemes in a time window from 390 to 425 ms after stimulus presentation ($p = 0.0046$). Both stimulus-locked effects were most pronounced over the bilateral centro-posterior electrodes. In the response-locked analysis, a length effect was identified as a difference between four and five phonemes from 335 to 320 ms before response onset, which was most pronounced over bilateral central and left posterior electrodes ($p = 0.0084$). Also, a length effect for the difference between four and six phonemes was revealed from 330 to 320 ms before response onset ($p = 0.0084$). This effect was most pronounced in right central and bilateral posterior electrodes.

Testing for a syllable frequency effect targeting phonetic encoding in the latency range from 400 to 450 ms after stimulus onset in the younger adults, the cluster-based permutation test revealed a difference between items with a high syllable frequency

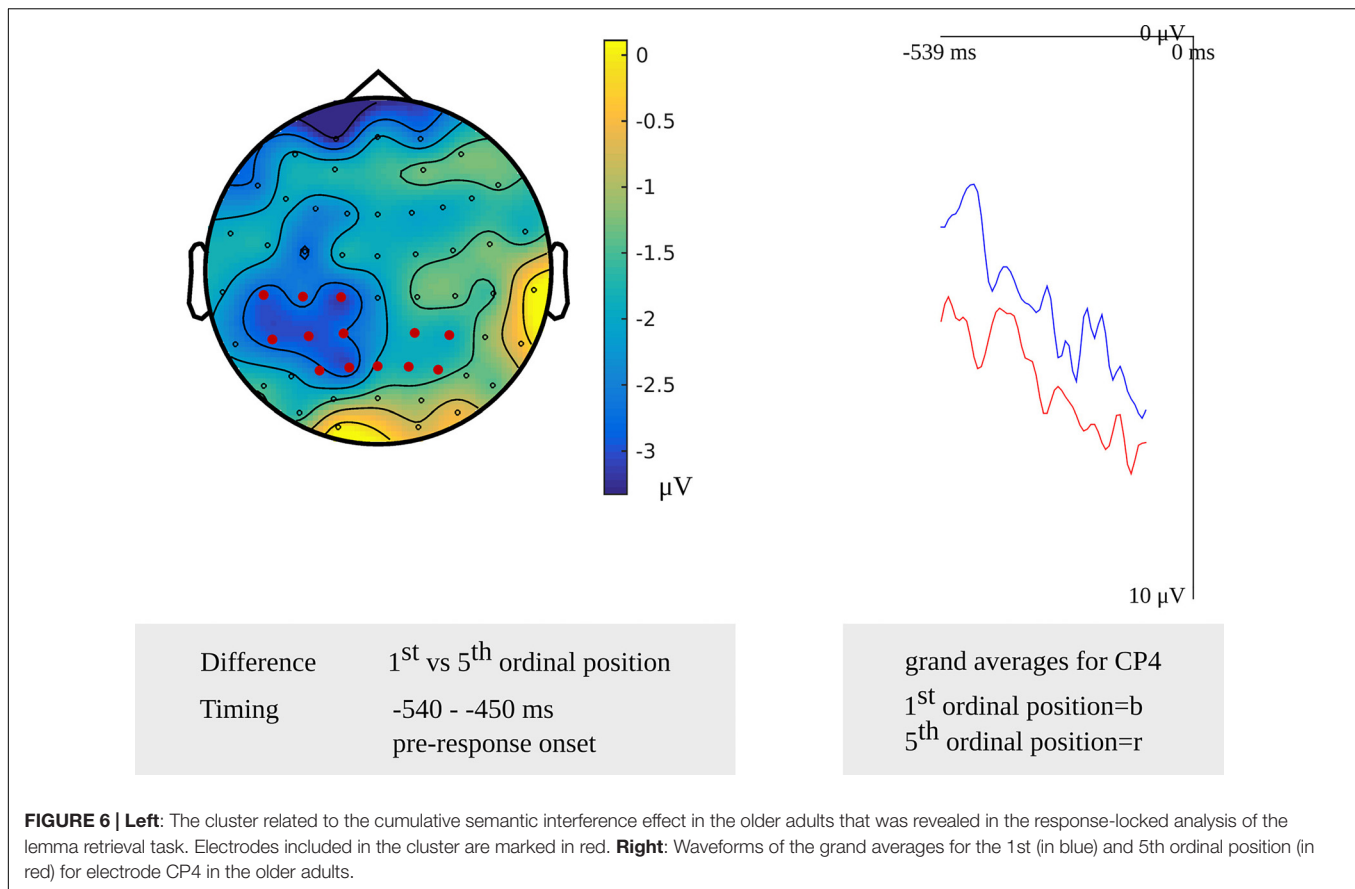
and items with a moderate syllable frequency ($p = 0.020$). In this latency range, the difference was most pronounced over the central sensors bilaterally. Another stimulus-locked syllable frequency effect was found as a difference between items with a high syllable frequency and items with a low syllable frequency in a time window from 350 to 450 ms after stimulus onset ($p = 0.012$), which is shown in **Figure 5**. The difference was most pronounced at the frontal and central sensors bilaterally. In **Figure 5**, the waveforms of the grand averages for the high and low syllable frequency items are provided as well. In the response-locked analysis, a difference between items with a high syllable frequency and items with a low syllable frequency was revealed in a time window from 250 to 200 ms before response onset ($p = 0.021$). The effect was most pronounced at bilateral central sensors.

Older Adults

In the older adults, testing for a cumulative semantic interference effect in the latency range from 540 to 450 ms before response onset, the cluster-based permutation test revealed a difference between the first and fifth ordinal positions ($p = 0.006$) that was taken as evidence for the stage of lemma retrieval. The difference was most pronounced over left posterior electrodes during the first 60 ms and most pronounced over the right posterior electrodes during the last 50 ms of the effect. No effect was found in the stimulus-locked analysis. The scalp distribution and the waveforms of the first and fifth ordinal position's grand average are shown in **Figure 6**.

For lexeme retrieval, an AoA effect was revealed in the cluster-based permutation analysis in three response-locked time windows as a difference between items with an early AoA (of around 5 years) and items with a moderate AoA (of around 6 years). The AoA effect was most pronounced over centro-posterior electrodes in the earliest cluster from 430 to 420 ms ($p = 0.012$) before response onset. In the second cluster, from 210 to 195 ms ($p = 0.009$) before response onset, the effect was most evident over the right frontal electrodes. The AoA effect was most distinct over right central electrodes in the last cluster with the longest duration from 165 to 140 ms ($p = 0.013$) before response onset, which is depicted in **Figure 7**. In **Figure 7**, the waveforms of the grand averages for the early and moderate AoA items are provided as well. No differences were found between items with an early AoA and items with a late AoA (of around 7 years). Also, no AoA effect was found in the stimulus-locked analysis.

For phonological encoding, the effect of the length in the number of phonemes on non-word reading was used in the cluster-based permutation analysis. In the older adults, a length effect was revealed as a difference between non-words with a length of four and six phonemes in the time windows from 100 to 135 ms ($p = 0.019$) and from 280 to 300 ms ($p = 0.0038$) after stimulus onset. In the first time window, the length effect was most pronounced over the right posterior electrodes, as shown in **Figure 8**. The waveforms of the grand averages for items consisting of four and six phonemes are provided in **Figure 8** as well. The effect was most pronounced over bilateral frontal and central electrodes in the second time window. No effects were



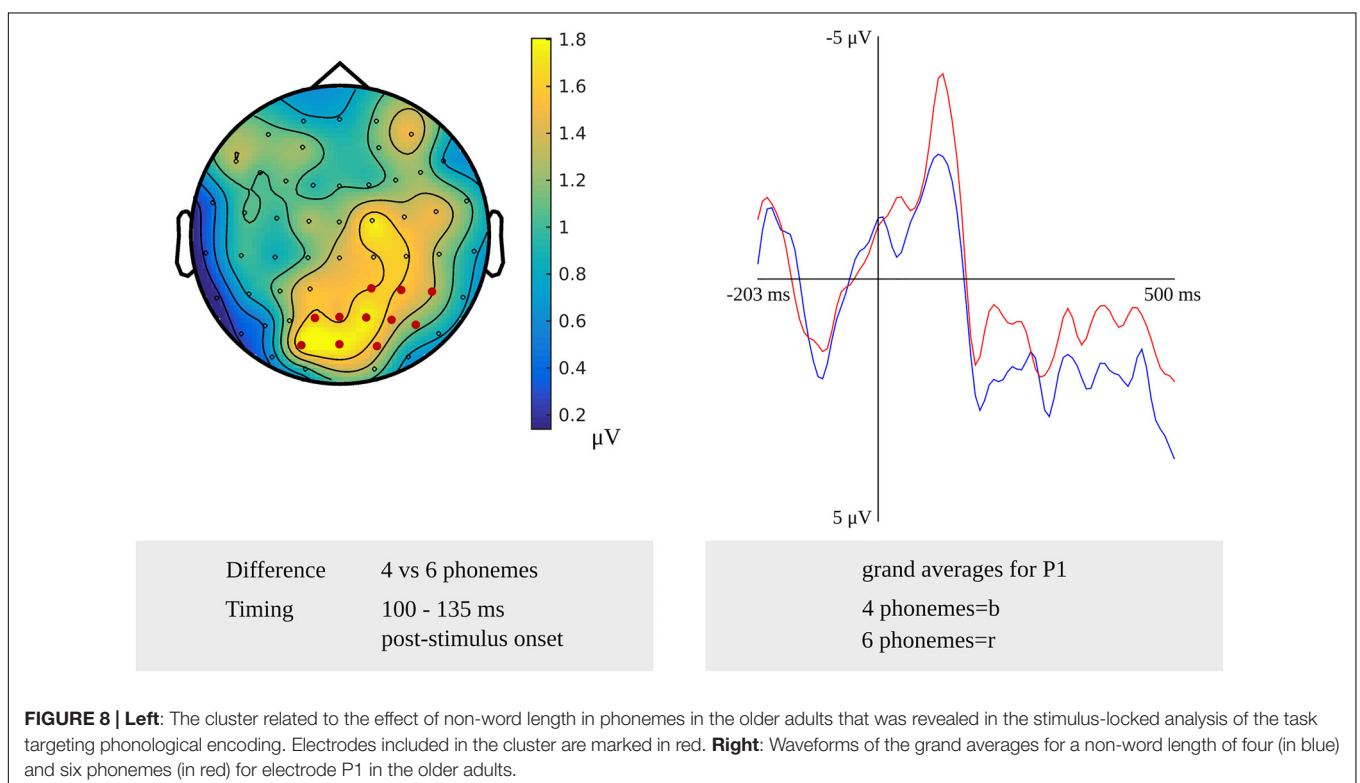
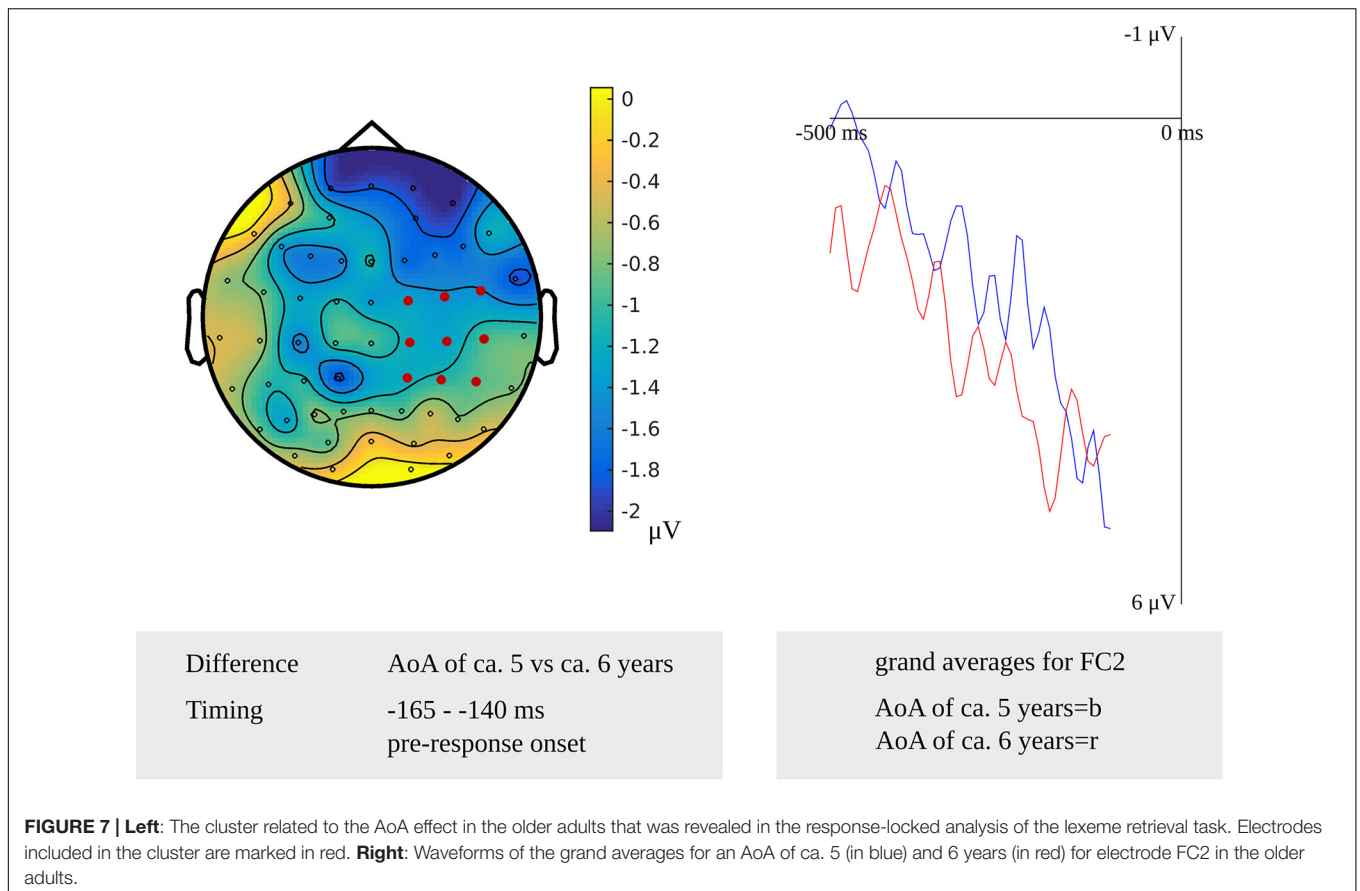
found for the comparison of non-words with a length of four and five phonemes. Also, no length effects were found in the response-locked analysis.

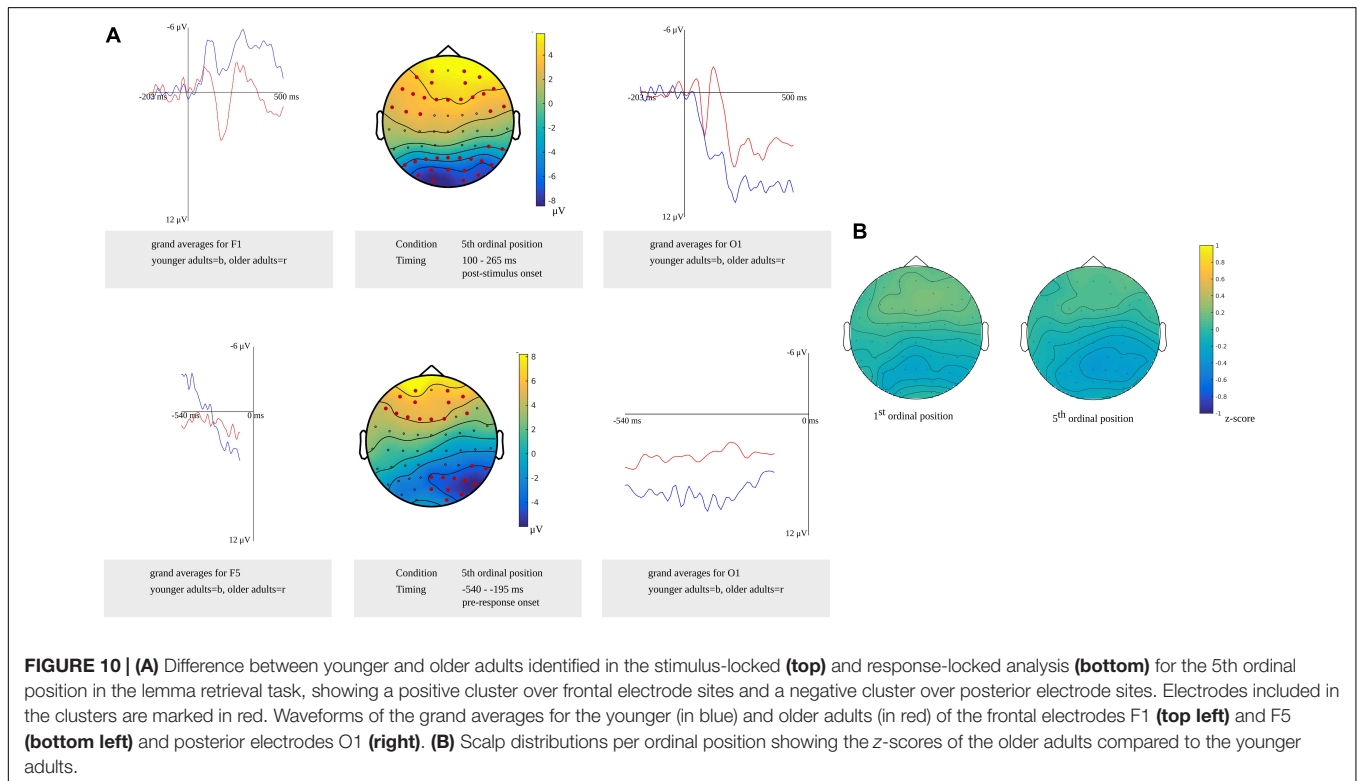
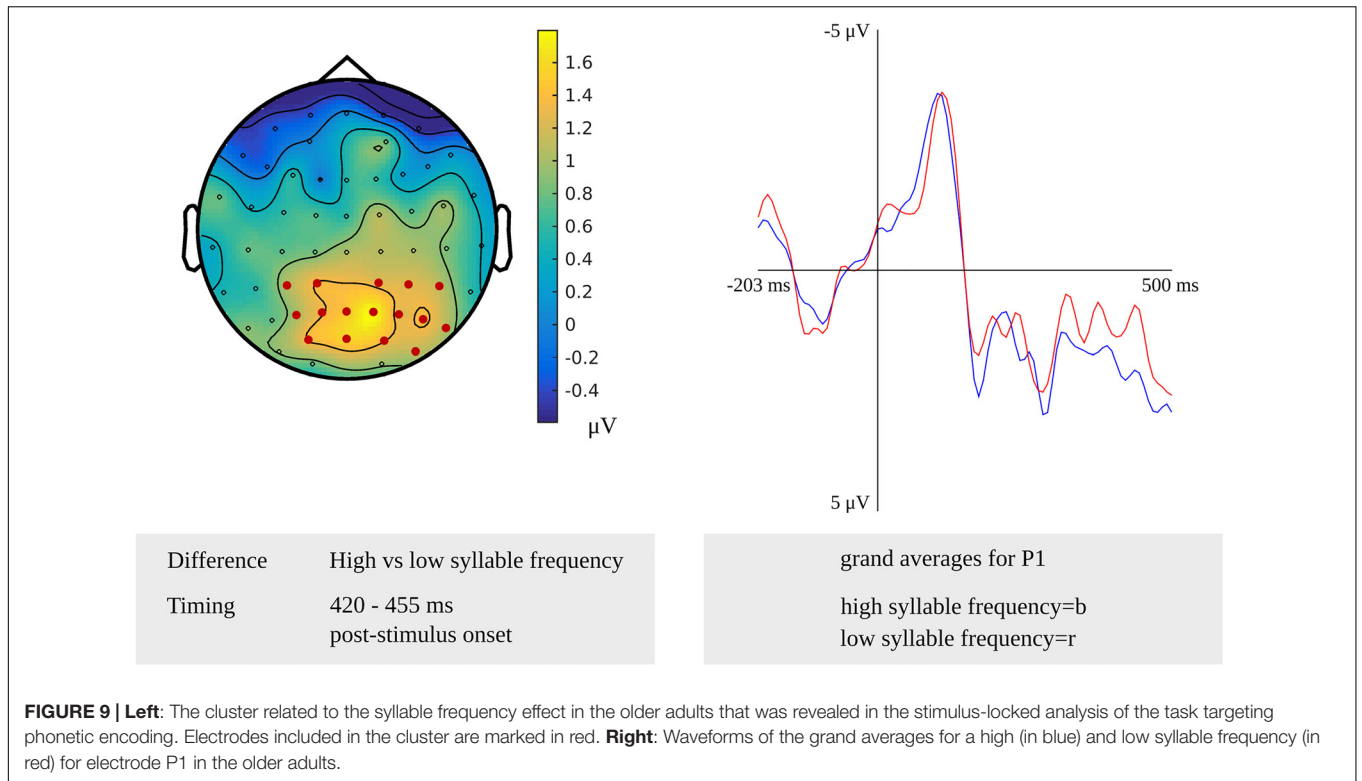
For tapping into phonetic encoding, the effect of syllable frequency on the non-word reading task was used. The stimulus-locked cluster-based permutation analysis revealed a syllable frequency effect for reading non-words with a high syllable frequency (ranging from 1,000 to 1,500) as compared to reading non-words with a moderate syllable frequency (ranging from 500 to 1,000) in a time window from 280 to 300 ms ($p = 0.0094$) and in a time window from 365 to 375 ms ($p = 0.022$) after stimulus presentation. The earliest effect was most pronounced over electrodes covering the right hemisphere, the later effect over the posterior electrodes. Furthermore, the comparison of non-words with a high syllable frequency to non-words with a low syllable frequency (ranging from 250 to 500) revealed effects from 280 to 290 ms ($p = 0.0196$) and from 420 to 455 ms ($p = 0.0078$) after stimulus onset. The effect starting at 280 ms was most pronounced over right-posterior electrodes, while the later effect shown in **Figure 9** was most pronounced over bilateral posterior electrodes. The waveforms of the high- and low-frequency items' grand averages are shown in **Figure 9** as well. Also, the syllable frequency effect was revealed from 455 to 435 ms ($p = 0.016$) before response onset. This effect was most pronounced over bilateral frontal and central electrodes.

Differences Between Younger and Older Adults

Comparing the older and younger adults in the time window for lemma retrieval in younger adults from 100 to 265 ms after stimulus presentation in the fifth ordinal position, the cluster-based permutation analysis showed that both groups differed. In this time window, two effects were identified: a positive ($p = 0.0026$) and a negative one ($p = 0.0022$). The electrodes over which the positive effect was most pronounced were located in frontal regions bilaterally. The negative effect was most pronounced in bilateral posterior regions. Also, in the time window for lemma retrieval in older adults from 540 to 450 ms before response onset, both groups were found to differ. Differences were observed as a positive ($p = 0.023$) effect that was most pronounced over bilateral frontal electrodes and a negative effect ($p = 0.013$) that was most pronounced over bilateral posterior electrodes. Furthermore, a difference between the groups was observed in the response-locked time window for lemma retrieval in the younger adults from 445 to 195 ms before response onset ($p = 0.0044$). This difference was most pronounced in the posterior regions bilaterally. The clusters are shown in **Figure 10A** along with the waveforms of the grand averages for younger and older adults.

Based on the results from the cluster-based permutation analysis, a time window from 540 to 450 ms before response onset in older adults was compared to a time window from 365 to 275 ms before response onset in young adults. The





z-scores computed for the first ($M = 0.03$, $SD = 0.15$, range = -0.37 to 0.27) and the fifth ordinal positions ($M = -0.12$; $SD = 0.15$, range = -0.41 to 0.19) indicated no differences in scalp distributions between the older and the younger adults. **Figure 10B** shows the *z*-scores of the individual electrodes mapped onto the scalp distribution per ordinal position.

In the time window for lexeme retrieval identified for the younger adults, from 100 to 300 ms after stimulus presentation, a difference between the older and younger adults was found for items with a moderate AoA ($p = 0.0022$). The difference was most pronounced in frontocentral regions bilaterally, as shown in **Figure 11A**. Also, the waveforms of the younger and older adults' grand averages are provided in **Figure 11A**. The response-locked time windows for lexeme retrieval from 430 and 140 ms before response onset identified in the older adults and from 475 to 330 ms before response onset identified in the younger adults did not reveal any differences between the groups.

The cluster-based permutation analysis targeting lexeme retrieval revealed no difference between early and late AoA conditions in the older adults; thus, the scalp distributions of the age groups could not be compared on these conditions. The age groups were compared on the early AoA and the moderate AoA conditions. A time window from 175 to 225 ms after stimulus presentation in the younger adults was compared to a time windows from 430 to 420 ms, from 210 to 195 ms, and from 165 to 140 ms before response onset in the older adults. Based on the *z*-scores of the electrodes, no differences in scalp distributions were found between the older and the younger adults for the early AoA ($M = 0.15$, $SD = 0.26$, range = -0.64 to 0.64) and the moderate AoA conditions ($M = 0.29$, $SD = 0.33$, range = -0.64 to 0.89). This is shown in **Figure 11B**.

The cluster-based permutation analysis for phonological encoding showed differences between older and younger adults for non-words consisting of five phonemes in a time window from 350 to 415 ms after stimulus presentation ($p = 0.015$). Also, for the non-words consisting of six phonemes, a difference between both age groups was found from 390 to 425 ms after stimulus presentation ($p = 0.014$). Both time windows were identified for phonological encoding in the young adults. The differences were most pronounced in bilateral posterior regions, as shown in **Figure 12A**. **Figure 12A** also shows the waveforms of the grand averages of the younger and the older adults. In the time windows identified for the older adults, no differences between the groups were found. This result was also the case for the response-locked time windows identified for phonological encoding in the younger adults.

For the older adults, no difference was found between non-words composed of four and five phonemes in the cluster-based analysis targeting phonological encoding, so the age groups cannot be compared on these conditions. The conditions with four and six phonemes were included in the scalp distributions analysis. Time windows from 390 to 425 ms after stimulus presentation and from 330 to 320 ms before response onset in the younger adults were compared to time windows from 105 to 135 ms and from 280 to 295 ms after stimulus presentation in the older adults. The *z*-scores revealed no differences in scalp distributions between the older and the younger adults for the

four phonemes condition ($M = -0.24$, $SD = 0.20$, range = -0.74 to 0.12) and the six phonemes condition ($M = -0.21$, $SD = 0.20$, range = -0.74 to 0.11). The scalp distributions are shown in **Figure 12B**.

For phonetic encoding, the cluster-based permutation analyses showed a difference between the older and the younger adults for moderate frequency non-words from 280 to 375 ms after stimulus presentation ($p = 0.007$). This range corresponds to the time window identified for phonetic encoding in the older adults. The groups did not differ in the time window for the younger adults. For low-frequency non-words, a difference between both groups was found from 280 to 455 ms after stimulus presentation ($p = 0.011$). This time window corresponds to the time window identified for phonetic encoding in older adults and also includes the time window in which phonetic encoding was identified in younger adults. Both effects were most pronounced in bilateral posterior regions, as shown in **Figure 13A**. This figure also shows the waveforms of the grand averages for the younger and older adults. No differences between the groups were found in the response-locked time windows.

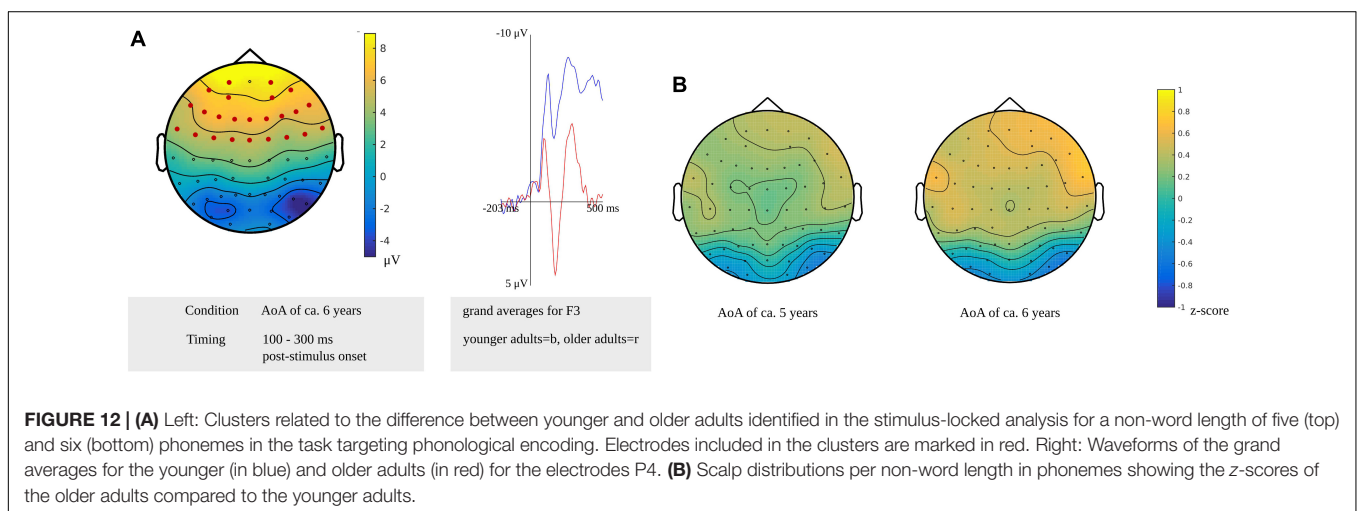
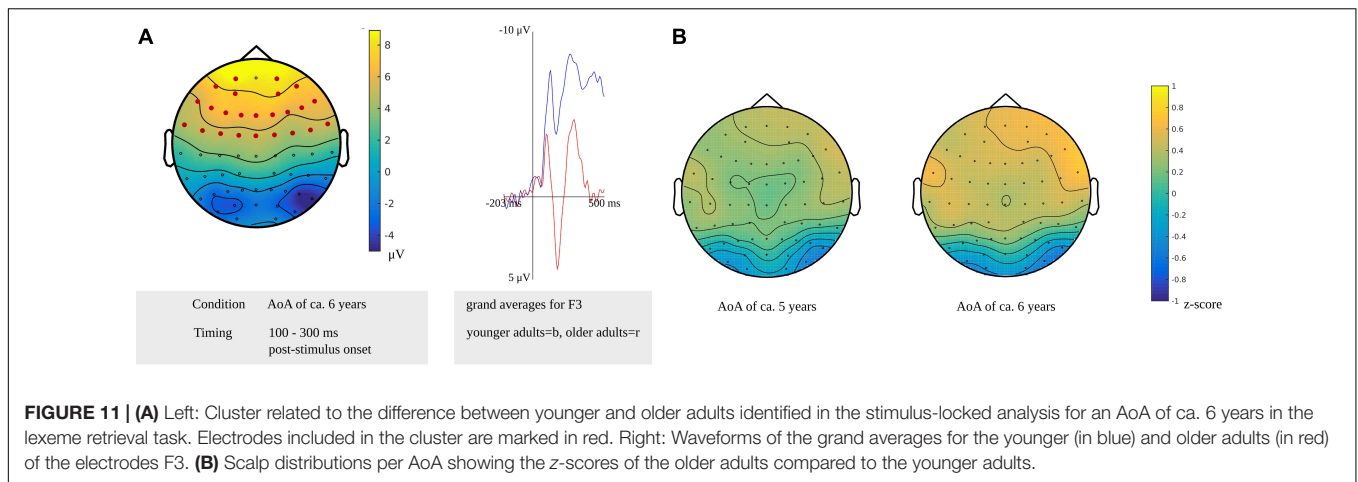
For non-words with a high syllable frequency and a moderate syllable frequency, a time window from 410 to 440 ms after stimulus presentation in younger adults was compared to time windows from 280 to 300 ms and from 365 to 375 ms after stimulus presentation in older adults. Based on the *z*-scores, no differences in scalp distributions were found between the older and the younger adults for both high frequency ($M = -0.15$, $SD = 0.11$, range = -0.33 to 0.10) and moderate frequency conditions ($M = -0.11$, $SD = 0.11$, range = -0.36 to 0.12). Also, *z*-scores for non-words with a high syllable frequency and a low syllable frequency were computed to compare a time window from 385 to 440 ms after stimulus presentation in younger adults to time windows from 280 to 290 ms and from 420 to 455 ms after stimulus presentation and from 450 to 460 ms before response onset in older adults. For the high-frequency ($M = -0.15$, $SD = 0.12$, range = -0.36 to 0.18) and the low-frequency conditions ($M = -0.11$, $SD = 0.14$, range = -0.44 to 0.17), no differences in scalp distributions based on the *z*-scores were found between older and younger adults. The scalp distributions are shown in **Figure 13B**.

DISCUSSION

The current study had two aims, which will be addressed in this discussion. The first was to identify the speech production stages in a group of older adults and in a group of younger adults. The second aim was to test whether the stages change with age with respect to the timing or regarding the neural configuration observed in the scalp distributions.

Identification of Speech Production Stages

To identify the stages of the speech production process, a protocol with EEG was developed with three tasks tapping into four speech production stages. The manipulations in the tasks used to identify the stages had an effect on the response times in



both the older and the younger adults. In the lemma retrieval task, the cumulative semantic interference effect caused increased response times for items belonging to the same category when they were presented at the fifth ordinal position compared to when they were presented at the first ordinal position. Also, later response times were found for items with a later AoA compared to items with an earlier AoA, as shown in the lexeme retrieval task. In the non-word reading task, non-words that consisted of more phonemes used to track phonological encoding and non-words with a lower syllable frequency used to tap into phonetic encoding caused increased response times. The results of the cluster-based permutation analysis of the EEG data revealed that the manipulations used in the tasks of the protocol showed an effect in particular time windows. First, the time windows in the younger adults will be discussed, after which the time windows in the older adults will be addressed.

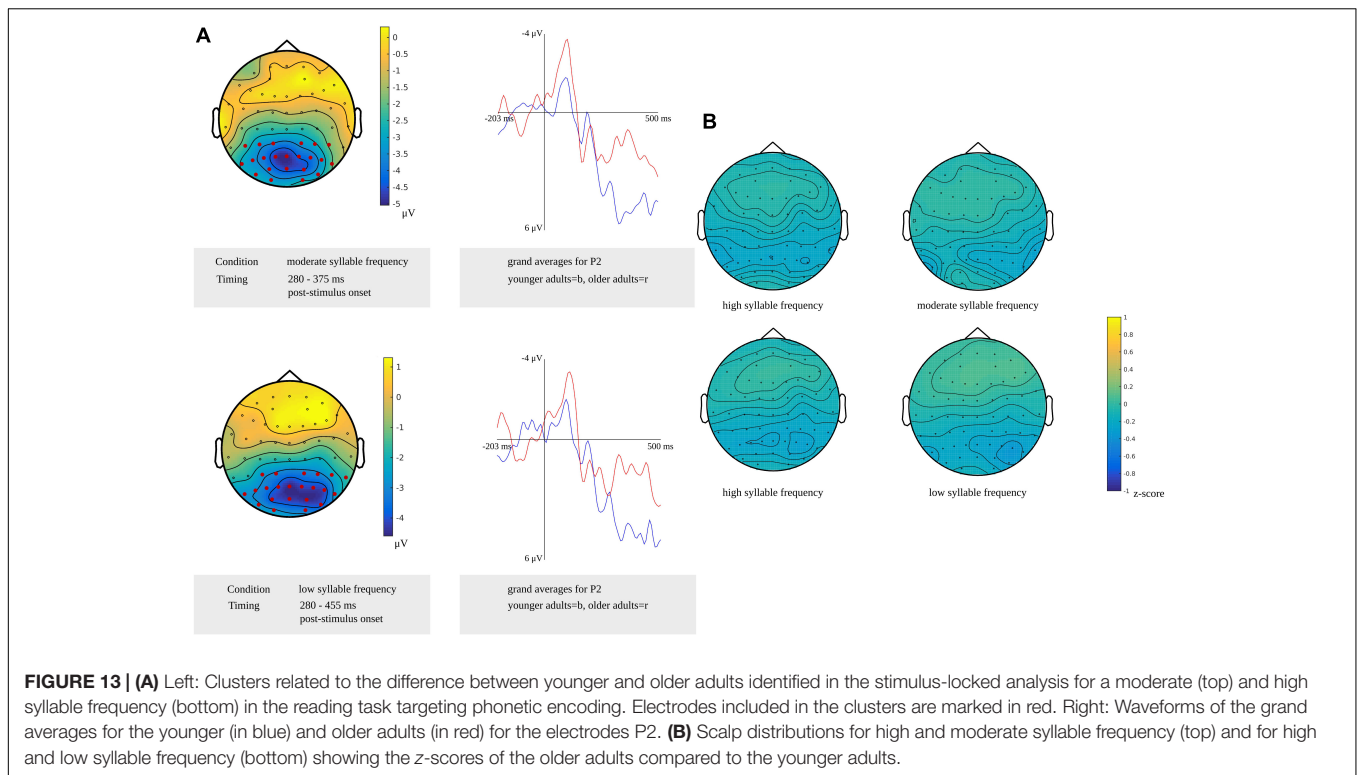
Younger Adults

In the younger adults, the timing of the cumulative semantic interference effect was revealed from 100 to 265 ms after stimulus presentation and from 445 to 195 ms before response onset. Response-locked cumulative semantic interference effects have

not been reported in previous studies using EEG. However, the stimulus-locked timing largely corresponded to the timing of this effect found by Maess et al. (2002) from 150 to 225 ms after stimulus presentation, but only partially overlapped with the timing of this effect found by Costa et al. (2009) from 200 to 380 ms after stimulus presentation. As our materials showed, the items used by Maess et al. (2002) depicted mono- and disyllabic high-frequency words. The materials used by Costa et al. (2009) also included longer and less-frequent words, which may explain the later latency of the cumulative semantic interference effect.

The timing of the AoA effect for the younger adults appeared from 100 to 300 ms after stimulus presentation. This result corresponds to the timing of this effect from 120 to 350 ms after stimulus presentation found by Laganaro and Perret (2011). Also, the response-locked effect for the younger adults from 475 to 330 ms before response onset overlaps with previously reported time windows of this stage from 380 after stimulus presentation up to 200 ms (Laganaro et al., 2012) or up to 100 ms before response onset (Valente et al., 2014).

Non-word length in phonemes was found to have an effect from 350 to 425 after stimulus presentation and from 335 to 320 before response onset for the younger adults. No previous



speech production studies using EEG have reported on non-word length effects. Word length effects have been studied using picture-naming tasks, but no effects have been identified (Valente et al., 2014; Hendrix et al., 2017). In our study, a length effect was identified with a non-word reading task. The input for phonological encoding of a word differs from the input for phonological encoding of a non-word, which may explain why the effect was found for non-words, but not for words. The phonological encoding of a familiar lexeme likely required less effort than the phonological encoding of an unfamiliar string of phonemes.

The syllable frequency effect in the non-word reading task has been identified after stimulus presentation from 350 to 450 ms for younger adults. Also, the effect has been found before response onset from 250 to 200 ms. Bürki et al. (2015), using syllable frequency effect in a non-word reading task, identified this effect from 170 to 100 ms before response onset. This effect was later than the effect found in the current study, most likely because the task required participants to insert a phoneme into the non-word as they read it, which complicated the task.

The time windows described in the previous paragraphs correspond to the speech production stages identified by Levelt et al. (1999) and Indefrey (2011). In the speech production model, lemma retrieval precedes lexeme retrieval. In the younger adults, the cumulative semantic interference effect and the AoA effect started at the same time in the stimulus-locked analysis, but the AoA effect lasted longer than the cumulative semantic interference effect. In the response-locked analysis, the cumulative semantic interference effect lasted longer than the AoA effect. The time window for lexeme retrieval started before

and ended during the time window for lemma retrieval. In the lexeme retrieval task, lemma retrieval was not manipulated, and thus, lemma retrieval was less demanding (and, hence, faster) in the lexeme retrieval task than in the lemma retrieval task. Therefore, the time window for lexeme retrieval in the lexeme retrieval task may have started earlier than the time window for lemma retrieval in the lemma retrieval task.

Lexeme retrieval is followed by phonological encoding in the model. For picture naming, the lexical route is used, whereas for non-word reading, the sublexical route should be recruited. Thus, the timing of the lexeme retrieval stage in the picture-naming task and the timing of the phonological encoding stage in the non-word reading task cannot be compared using our method. Phonological encoding precedes phonetic encoding in the model. In the stimulus-locked analysis, the non-word length effect started at the same time as the syllable frequency effect, but the length effect ended earlier. In the response-locked analysis, the non-word length in phonemes effect preceded the syllable frequency effect. Thus, the protocol can be used to identify the stages using EEG in the younger adults.

Older Adults

In the older adults, the cumulative semantic interference effect was found from 540 to 450 ms before response onset. Since no response-locked cumulative semantic interference effects have been reported previously, the response-locked effect revealed in the older adults cannot be compared to other studies.

AoA effects have previously been identified in response-locked time windows until 200 ms (Laganaro et al., 2012) or 100 ms before response onset (Valente et al., 2014). These time windows

overlap with the response-locked effects for the older adults from 430 to 140 ms before response onset.

The effect of non-word length in phonemes was identified from 100 to 135 ms and from 280 to 300 ms after stimulus presentation for the older adults. This study is the first to report the effects of non-word length in number of phonemes in an EEG study.

The second effect that was tested in the non-word reading task was syllable frequency, which has been identified from 280 to 455 ms after stimulus presentation. This effect was found from 455 to 435 ms before response onset as well. The timing of these effects is earlier than the timing of the syllable frequency effect reported by Bürki et al. (2015). As said above, task was more demanding, which may explain these differences.

In the older adults, the response-locked cumulative semantic interference effect preceded the response-locked AoA effect. This corresponds to the speech production processes identified by Levelt et al. (1999), Indefrey (2011), in which lemma retrieval precedes lexeme retrieval. In the older adults, the effect of non-word length in phonemes was identified before the syllable frequency effect, but there is an overlap of 20 ms in the stimulus-locked analysis. This finding is also in agreement with the model, because phonological encoding precedes phonetic encoding. Thus, the protocol can be used to identify the stages using EEG in the older adults as well.

Aging Effects on Speech Production Stages

The behavioral data showed that both the younger adults and the older adults performed at ceiling on every task. Thus, in contrast to the study by Connor et al. (2004), no reduced accuracy in picture naming was found for older adults. This can be explained by a major difference in the age range of the participants in both studies: it was larger in the study by Connor et al. (2004: from 30 to 94 years) than in the current study, from 17 to 65 years. A behavioral difference between the groups was found in the response times. The older adults responded later than the younger adults on every task. It was hypothesized that the later response times of the older adults should reflected in the timing of the speech production stages in the EEG.

Differences in Timing Between Younger and Older Adults

Lemma retrieval requires semantic memory to activate the target lemma node along with its semantically related neighbors. These neighbors are inhibited to select the target lemma. Since both semantic memory (Cardenas et al., 2011; Harada et al., 2013) and inhibition (Harada et al., 2013) decline with aging, the duration of the lemma retrieval stage was expected to be increased in older adults. This hypothesis was not confirmed, because the lemma retrieval stage lasted 90 ms in the older adults, while in the younger adults, its duration was 165 ms in the stimulus-locked analysis and 250 ms in the response-locked analysis. However, all time windows of the effects that were found in the older adults were shorter than the time windows of the effects found in the younger adults. In older adults, neurons that fire together are possibly less synchronous in their timing, less aligned regarding

their geometry, or the effect has a more variable latency (Wlotko et al., 2010). Therefore, the time window in which all participants show an effect is shorter.

Since the duration of lemma retrieval was expected to be increased, the onset of the next stage, lexeme retrieval, was expected to be delayed in the older adults. This hypothesis was confirmed. The response-locked effect started 45 ms later for the older adults compared to the younger adults. Also, an increased duration of the lexeme retrieval stage was hypothesized, because of the tip-of-the-tongue phenomenon, which is observed more frequently in older adults (Shafto et al., 2007). No increased duration was found, which again can be explained by the reduction in the effect caused by the effect's variability within and between the older adults (Wlotko et al., 2010).

The stages of the sublexical route were expected not to be delayed in older adults. There have been no previous studies on aging's effect on phonological encoding. Also, older adults have not revealed longer response times producing alternating syllable strings, which require more effort during phonetic encoding, than for the production of sequential syllable strings (Tremblay and Deschamps, 2016). However, both the effect of non-word length in phonemes related to phonological encoding and the syllable frequency effect targeting phonetic encoding started earlier for the older adults than for the younger adults. The difference in the onset of the timing of these stages between the groups is quite large; hence, this difference cannot be explained by the effect's variability in older adults.

Neurophysiological Differences Between Younger and Older Adults

There were differences between the younger and the older adults regarding the time windows in which effects that were related to the stages were found. Results of the cluster-based permutation analyses showed that for every stage in at least one time window, differences between younger and older adults were found. In the time windows in which the younger adults showed a cumulative semantic interference effect, an AoA effect, or an effect of non-word length in number of phonemes, no such effect was observed in the older adults. This finding shows that the older adults had a different timing for the speech production stages than the younger adults. Despite partially overlapping time windows for the syllable frequency effect in the younger and older adults, a difference between both groups was found. The overlap in timing was possibly too short, so both groups differed during the majority of the time window, or the neural configuration of the syllable frequency effect differed between the groups. Except for the response-locked time windows identified using the cumulative semantic interference effect, differences between younger and older adults were generally identified in stimulus-locked time windows. When the stimulus is presented, the first process is the visual analysis of the picture or the non-word. This process is assumed to be identical in both age groups, because the efficiency of the visual network is not expected to change with age (Geerligts et al., 2015). After that, higher cognitive function networks, such as CON and FPCN are involved in the speech production stages. A decrease in the local efficiency of these networks may alter their neural signature or change their timing,

which is reflected in the EEG. Even though the older participants in the study by Geerligts et al. were, on average, almost a decade older than the older adults in our study, our older participants may have a mild decrease in local efficiency and modularity in the CON and the FPCN compared to the younger adults, because the decrease is not linear with age (Geerligts et al., 2015).

An overview of the timing of the stages in the younger and older adults and the timing of significant differences between the two groups is provided in **Figure 14**.

Apart from the timing of the speech production stages, the neural configurations of the scalp distributions of the stages have been compared between the older and the younger adults. It was hypothesized that the scalp distributions do not change with age, because the same groups of neurons are expected to be involved in the stages of speech production in neurologically healthy adults, regardless of the adults' age. Despite the fact that the effects related to each stage have been found in different time windows in the two groups, the scalp distributions during the stage were identical in the older and younger adults. This uniformity was the case for each speech production stage. Therefore, it can be

concluded that older adults used the same neuronal processes as younger adults in the speech production stages. This was also supported by our behavioral results. Like the younger adults, the older adults performed at ceiling on the tasks. Also, the response times showed that the manipulations used in the tasks had the same effects in older and younger adults. Thus, the same factors had an influence on the speech production stages in both age groups.

The question remains why the response times of the older adults were later than the response times of the younger adults, even though the timing of the effects used to target the speech production stages was not generally delayed in the older adults. In the lexical route, lexeme retrieval was found to be delayed in older compared to younger adults. Since both picture-naming tasks required lexeme retrieval, the delay before this stage may have resulted in longer response times on the lemma and lexeme retrieval tasks. This is in line with the findings in the study by Laganaro et al. (2012) revealing differences between slow and fast speakers before the time window in which the AoA effect was found.

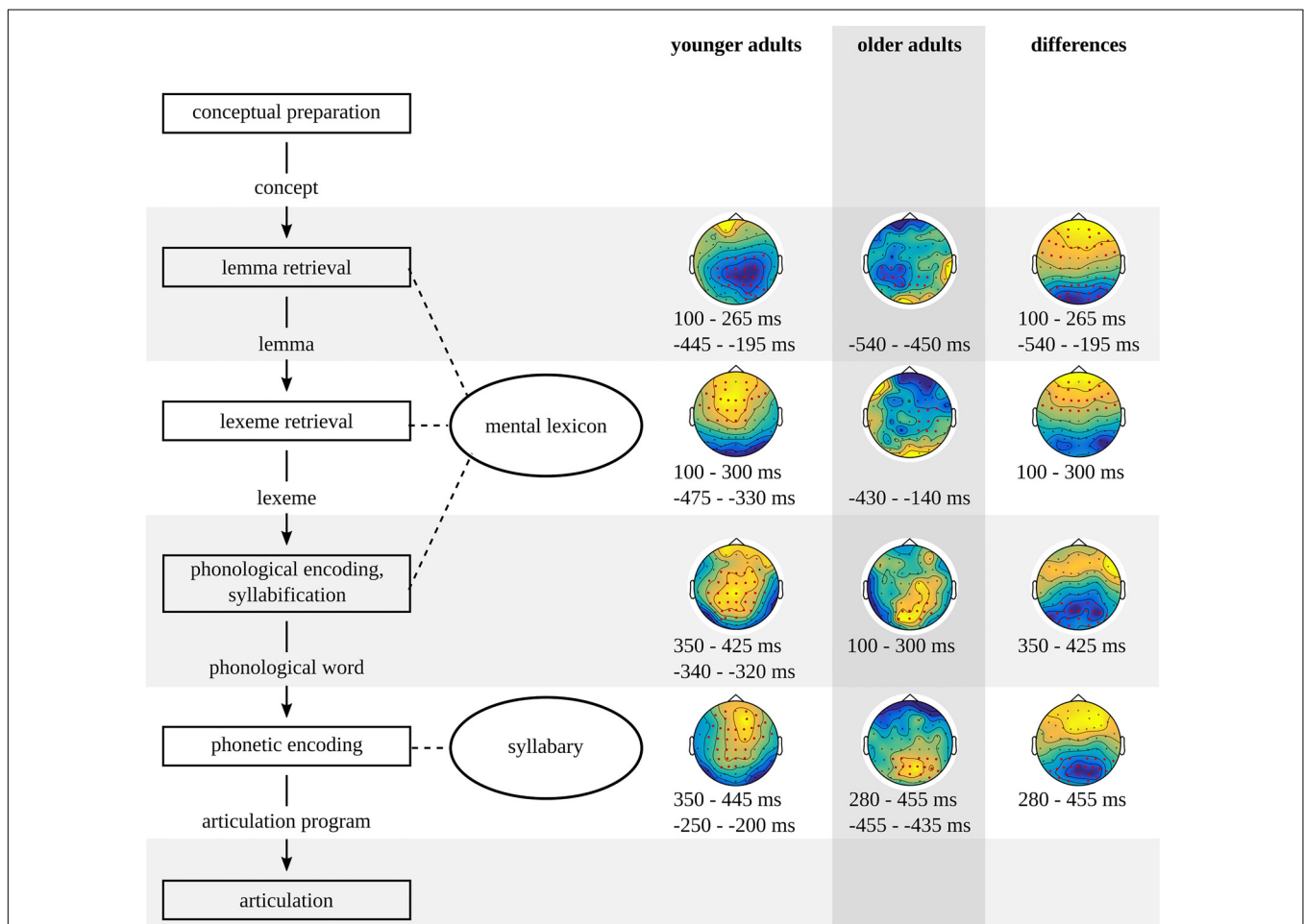


FIGURE 14 | Timing of the stages in the model of spoken word and non-word production based on the results of the younger and the older adults and their differences.

Lexeme retrieval is not involved in non-word production. Therefore, delayed lexeme retrieval cannot explain later response times on non-word tasks in older adults, while no delay was observed for the phonological and phonetic encoding stages. Maybe, older adults respond later, because they generally are slower, as suggested in the Global Slowing Hypothesis (e.g., Brinley, 1965). However, this should have been reflected in the EEG as a longer duration and a later onset for every speech production stage, because neurophysiological measures are more sensitive than response time measures. Participants were asked to name the items as fast and accurately as possible. The tasks were fairly easy, so the accuracy of all patients was at ceiling. While younger adults can respond fast and accurately at the same time, older adults are known to focus on either speed or accuracy (Ratcliff et al., 2007). Maybe older adults focused more on accuracy in our study and, therefore, needed to collect more information before they were ready to respond (Rabbitt, 1979). In that case, the processes may not have been delayed in general, but only the decision whether the response was accurate or not was delayed. Thus, after the speech production process has been planned to its final stage, articulation, the older adults may have waited longer than the younger adults until they responded. In that case, this effect is not visible on the EEG, but only reflected in longer response times. If older adults wait before responding, the response-locked effects should be identified earlier in the older adults than in the younger adults. This, indeed, was the case for the cumulative semantic interference effect and the syllable frequency effect, but not for the AoA effect. However, individual differences are known to modulate the time window of the AoA effect (Laganaro et al., 2012). A possible modulation of the AoA effect is supported by our response time data, in which the older adults showed a smaller AoA effect than the younger adults.

CONCLUSION

To conclude, the stages of the speech production process have been successfully identified in older and younger adults using the tasks of the protocol with EEG. The manipulations in the tasks had the same effect on the response time in both age groups; thus, the same factors influenced the speech production stages. Also, the scalp distributions related to the speech production stages did not differ between the older and the younger adults. This shows that the same neural processes are used during the speech production stages.

However, behaviorally, the comparison of the older and the younger adults showed that the older adults required longer response times on all tasks. Yet, the EEG results showed that the speech production stages do not generally start later or last longer in the older adults compared to the younger adults.

LIMITATIONS

The study is subject to two potential limitations. In this study, we included older adults (40–65 years old), whereas it is common practice to compare younger adults (i.e., university students) to a group of elderly (usually over 70 years old). Thus, the age difference between the younger and older adults was smaller than in other studies that compare language production and, therefore, the aging effects found in the current study are potentially not as large as when younger and individuals with aphasia is now possible: individuals with aphasia and without concomitant cognitive disorders are usually within the age range of our group of older adults. However, it would be very interesting to compare the performance of both age groups of the current study with the healthy elderly and individuals with dementia, who are usually above 70 years old.

Second, non-word reading skills of the two groups included in the present study have not been assessed prior to the experiment. Reading was only assessed using self-report, which cannot be used to detect potential variation in reading skills. This potential variation may have had an effect at the phonological and phonetic encoding stages. We do not think this caveat influenced the results, however, because all participants performed at ceiling on the non-word reading task.

ETHICS STATEMENT

This study was approved by the Research Ethics Committee of the Faculty of Arts of the University of Groningen.

AUTHOR CONTRIBUTIONS

JH is working on this Ph.D. project, did the actual studies, and wrote the largest part of the text. RB is promotor and PI of this project, and wrote a large part of the manuscript. RJ is daily supervisor of JH. PM initiated this project.

FUNDING

This research was supported by an Erasmus Mundus Joint Doctorate (EMJD) Fellowship for “International Doctorate for Experimental Approaches to Language And Brain” (IDEALAB) of the University of Groningen (Netherlands), University of Newcastle (United Kingdom), University of Potsdam (Germany), University of Trento (Italy), and Macquarie University, Sydney (Australia), under Framework Partnership Agreement 2012-0025, specific grant agreement number 2015-1603/001-001-EMJD, awarded to JH by the European Commission. RB is partially supported by the Center for Language and Brain, National Research University Higher School of Economics, RF Government grant, agreement number 14.641.31.0004.

REFERENCES

- Baayen, H. R., Piepenbrock, R., and Gulikers, L. (1995). *The CELEX Lexical Database. Linguistic Data Consortium*. Philadelphia, PA: University of Pennsylvania.
- Bastiaanse, R. (2010). *Auditiief Taalbegri Psprogramma (ATP)*. Houten: Bohn Stafleu van Loghum.
- Bastiaanse, R., Wieling, M., and Wolthuis, N. (2016). The role of frequency in the retrieval of nouns and verbs in aphasia. *Aphasiology* 30, 1221–1239. doi: 10.1080/02687038.2015.1100709
- Boersma, P., and Weenink, D. (2018). *Praat: Doing Phonetics by Computer*. Available at: <http://www.praat.org/>
- Brinley, J. F. (1965). "Cognitive sets, speed and accuracy of performance in the elderly," in *Behavior, Aging, and the Nervous System*, eds A. T. Welford and J. E. Birren (Springfield, IL: Thomas), 114–149.
- Brybaert, M., Stevens, M., De Deyne, S., Voorspoels, W., and Storms, G. (2014). Norms of age of acquisition and concreteness for 30,000 Dutch words. *Acta Psychol.* 150, 80–84. doi: 10.1016/j.actpsy.2014.04.010
- Bürki, A., Pellet-Cheneval, P., and Laganaro, M. (2015). Do speakers have access to a mental syllabary? ERP comparison of high frequency and novel syllable production. *Brain Lang.* 150, 90–102. doi: 10.1016/j.bandl.2015.08.006
- Cardenas, V. A., Chao, L. L., Studholme, C., Yaffe, K., Miller, B. L., Madison, C., et al. (2011). Brain atrophy associated with baseline and longitudinal measures of cognition. *Neurobiol. Aging* 32, 572–580. doi: 10.1016/j.neurobiolaging.2009.04.011
- Connor, L. T., Spiro, A., Obler, L. K., and Albert, M. L. (2004). Change in object naming ability during adulthood. *J. Gerontol. Ser. B.* 59, 203–209. doi: 10.1093/geronb/59.5.P203
- Costa, A., Strijkers, K., Martin, C., and Thierry, G. (2009). The time course of word retrieval revealed by event-related brain potentials during overt speech. *Proc. Natl. Acad. Sci. U.S.A.* 106, 21442–21446. doi: 10.1073/pnas.0908921106
- Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009
- Dosenbach, N. U. F., Fair, D. A., Miezin, F. M., Cohen, A. L., Wenger, K. K., Dosenbach, R. A. T., et al. (2007). Distinct brain networks for adaptive and stable task control in humans. *Proc. Natl. Acad. Sci. U.S.A.* 104, 11073–11078. doi: 10.1073/pnas.0704320104
- E-Prime 2.0 (2012). *Psychology Software Tools*. Sharpsburg, PA: Psychology Software Tools, Inc.
- Fjell, A. M., and Walhovd, K. B. (2011). Structural brain changes in aging: courses, causes and cognitive consequences. *Rev. Neurosci.* 21:187.
- Forstmann, B. U., Tittgemeyer, M., Wagenmakers, E.-J., Derrfuss, J., Imperati, D., and Brown, S. (2011). The speed–accuracy tradeoff in the elderly brain: a structural model-based approach. *J. Neurosci.* 31, 17242–17249. doi: 10.1523/JNEUROSCI.0309-11.2011
- Freeman, S. H., Kandel, R., Cruz, L., Rozkalne, A., Newell, K., Frosch, M. P., et al. (2008). Preservation of neuronal number despite age-related cortical brain atrophy in elderly subjects without Alzheimer disease. *J. Neuropathol. Exp. Neurol.* 67, 1205–1212. doi: 10.1097/NEN.0b013e31818f72f
- Geerligs, L., Renken, R. J., Saliassi, E., Maurits, N. M., and Lorist, M. M. (2015). A brain-wide study of age-related changes in functional connectivity. *Cereb. Cortex* 25, 1987–1999. doi: 10.1093/cercor/bhu012
- Harada, C. N., Natelson Love, M. C., and Triebel, K. L. (2013). Normal cognitive aging. *Clin. Geriatr. Med.* 29, 737–752. doi: 10.1016/j.cger.2013.07.002
- Hendrix, P., Bolger, P., and Baayen, H. (2017). Distinct ERP signatures of word frequency, phrase frequency, and prototypicality in speech production. *J. Exp. Psychol.* 43, 128–173. doi: 10.1037/s00400332
- Nederlandse Taalunie (2004). *Corpus Gesproken Nederlands, Version 2.0*. Leiden: TST-Centrale INL.
- Howard, D., Nickels, L., Coltheart, M., and Cole-Virtue, J. (2006). Cumulative semantic inhibition in picture naming: experimental and computational studies. *Cognition* 100, 464–482. doi: 10.1016/j.cognition.2005.02.006
- Indefrey, P. (2011). The spatial and temporal signatures of word production components: a critical update. *Front. Psychol.* 2:225. doi: 10.3389/fpsyg.2011.00255
- Jasper, H. H. (1958). Report of the committee on methods of clinical examination in electroencephalography. *Electroencephalogr. Clin. Neurophysiol.* 10, 370–375. doi: 10.1016/0013-4694(58)90053-1
- Laganaro, M., and Perret, C. (2011). Comparing electrophysiological correlates of word production in immediate and delayed naming through the analysis of word age of acquisition effects. *Brain Topogr.* 24, 19–29. doi: 10.1007/s10548-010-0162-x
- Laganaro, M., Valente, A., and Perret, C. (2012). Time course of word production in fast and slow speakers: a high density ERP topographic study. *Neuroimage* 59, 3881–3888. doi: 10.1016/j.neuroimage.2011.10.082
- Levelt, W. J., Roelofs, A., and Meyer, A. S. (1999). A theory of lexical access in speech production. *Behav. Brain Sci.* 22, 38–75.
- Levelt, W. J. M., and Wheeldon, L. (1994). Do speakers have access to a mental syllabary? *Cognition* 50, 239–269. doi: 10.1016/0010-0277(94)90030-2
- Maess, B., Friederici, A. D., Damian, M., Meyer, A. S., and Levelt, W. J. M. (2002). Semantic category interference in overt picture naming: sharpening current density localization by PCA. *J. Cogn. Neurosci.* 14, 455–462. doi: 10.1162/089892902317361967
- Maris, E., and Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* 164, 177–190. doi: 10.1016/j.jneumeth.2007.03.024
- Marnier, L., Nyengaard, J. R., Tang, Y., and Pakkenberg, B. (2003). Marked loss of myelinated nerve fibers in the human brain with age. *J. Comp. Neurol.* 462, 144–152. doi: 10.1002/cne.10714
- MATLAB (2015). *MATLAB*. Natick, MA: The MathWorks Inc.
- Nieuwenhuis, S., Forstmann, B. U., and Wagenmakers, E.-J. (2011). Erroneous analyses of interactions in neuroscience: a problem of significance. *Nat. Neurosci.* 14, 1105–1107. doi: 10.1038/nn.2886
- 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
- Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J.-M. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci.* 2011, 1–9. doi: 10.1155/2011/156869
- R Core Team (2017). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Rabbitt, P. (1979). How old and young subjects monitor and control responses for accuracy and speed. *Br. J. Psychol.* 70, 305–311. doi: 10.1111/j.2044-8295.1979.tb01687.x
- Ratcliff, R., Thapar, A., and McKoon, G. (2007). Application of the diffusion model to two-choice tasks for adults 75–90 years old. *Psychol. Aging* 22, 56–66. doi: 10.1037/0882-7974.22.1.56
- Sahin, N. T., Pinker, S., Cash, S. S., Schomer, D., and Halgren, E. (2009). Sequential processing of lexical, grammatical, and phonological information within Broca's area. *Science* 326, 445–449. doi: 10.1126/science.1174481
- Shafto, M. A., Burke, D. M., Stamatakis, E. A., Tam, P. P., and Tyler, L. K. (2007). On the tip-of-the-tongue: neural correlates of increased word-finding failures in normal aging. *J. Cogn. Neurosci.* 19, 2060–2070. doi: 10.1162/jocn.2007.19.12.2060
- Thatcher, R. W., Biver, C. J., and North, D. M. (2002). *Z Score EEG Biofeedback: Technical Foundations*. Seminole, FL: Applied Neuroscience Inc.
- Tremblay, P., and Deschamps, I. (2016). Structural brain aging and speech production: a surface-based brain morphometry study. *Brain Struct. Funct.* 221, 3275–3299. doi: 10.1007/s00429-015-1100-1
- Valente, A., Bürki, A., and Laganaro, M. (2014). ERP correlates of word production predictors in picture naming: a trial by trial multiple regression analysis from stimulus onset to response. *Front. Neurosci.* 8:390. doi: 10.3389/fnins.2014.00390
- Vincent, J. L., Kahn, I., Snyder, A. Z., Raichle, M. E., and Buckner, R. L. (2008). Evidence for a frontoparietal control system revealed by intrinsic

- functional connectivity. *J. Neurophysiol.* 100, 3328–3342. doi: 10.1152/jn.90355.2008
- Wlotko, E. W., Lee, C. L., and Federmeier, K. D. (2010). Language of the aging brain: event-related potential studies of comprehension in older adults: language of the aging brain. *Lang. Linguist. Compass* 4, 623–638. doi: 10.1111/j.1749-818X.2010.00224.x
- Zheng, F., Liu, Y., Yuan, Z., Gao, X., He, Y., Liu, X., et al. (2018). Age-related changes in cortical and subcortical structures of healthy adult brains: a surface-based morphometry study: age-related study in healthy adult brain structure. *J. Magn. Reson. Imaging* 49, 152–163. doi: 10.1002/jmri.26037

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 den Hollander, Jonkers, Mariën and Bastiaanse. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

APPENDIX 1A | EEG statistics for the younger adults.

Analysis	Comparison	Time domain	Probability	Cluster statistics	Standard deviation	Confidence interval range
Lemma retrieval						
Stimulus locked	1st vs. 5th ordinal position	100 to 265 ms	<0.001	-1,505.0	<0.001	<0.001
Response locked	1st vs. 5th ordinal position	-445 to -195 ms	0.005	-2,836.6	<0.001	0.002
Lexeme retrieval						
Stimulus locked	AoA ca. 5 years vs. ca. 6 years	100 to 300 ms	0.002	1,116.1	<0.001	0.001
Response locked	AoA ca. 5 years vs. ca. 7 years	-475 to -330 ms	<0.001	-1,954.7	<0.001	<0.001
Phonological encoding in reading						
Stimulus locked	Length 4 vs. 5 phonemes	350 to 415 ms	0.003	665.8	<0.001	0.002
	Length 4 vs. 6 phonemes	390 to 425 ms	0.005	317.9	<0.001	0.002
Response locked	Length 4 vs. 5 phonemes	-335 to -320 ms	0.008	200.7	0.001	0.002
	Length 4 vs. 6 phonemes	-330 to -320 ms	0.008	117.0	0.001	0.002
Phonetic encoding in reading						
Stimulus locked	High vs. moderate frequency	400 to 450 ms	0.020	316.5	0.002	0.004
	High vs. low frequency	350 to 450 ms	0.012	665.4	0.002	0.003
Response locked	High vs. low frequency	-250 to -200 ms	0.021	214.7	0.002	0.004

APPENDIX 1B | EEG statistics for the older adults.

Analysis	Comparison	Time window	Probability	Cluster statistics	Standard deviation	Confidence interval range
Lemma retrieval						
Response locked	1st vs. 5th ordinal position	-540 to -450 ms	0.006	-340.9	0.004	0.007
Lexeme retrieval						
Response locked	AoA ca. 5 vs. ca. 6 years	-430 to -420 ms	0.012	-78.8	0.002	0.003
		-210 to -195 ms	0.009	-96.7	0.001	0.003
		-165 to -140 ms	0.013	-131.6	0.002	0.003
Phonological encoding in reading						
Stimulus locked	Length 4 vs. 6 phonemes	100 to 135 ms	0.020	124.7	0.002	0.004
		280 to 300 ms	0.004	186.8	<0.001	0.002
Phonetic encoding in reading						
Stimulus locked	High vs. moderate frequency	280 to 300 ms	0.009	142.7	0.001	0.003
		365 to 375 ms	0.022	46.5	0.002	0.004
	High vs. low frequency	280 to 290 ms	0.020	59.2	0.002	0.004
		420 to 455 ms	0.008	174.2	0.001	0.002
Response locked	High vs. low frequency	-455 to -435 ms	0.016	98.6	0.002	0.004

APPENDIX 1C | EEG statistics for the comparison of the older and younger adults.

Analysis	Condition	Time window	Probability	Cluster statistics	Standard deviation	Confidence interval range
Lemma retrieval						
Stimulus locked	5th ordinal position	100 to 265 ms	0.003	907.6	<0.001	0.001
			0.002	-1,088.3	<0.001	0.001
Response locked	5th ordinal position	-540 to -450 ms	0.023	255.8	0.002	0.004
		-445 to -195 ms	0.013	-436.6	0.002	0.003
			0.004	-2,139.8	<0.001	0.002
Lexeme retrieval						
Stimulus locked	AoA ca. 6 years	100 to 300 ms	0.002	-1,749.4	<0.001	0.001
Phonological encoding in reading						
Stimulus locked	5 phonemes	350 to 415 ms	0.015	-386.5	0.002	0.003
		6 phonemes	0.014	-227.5	0.002	0.003
Phonetic encoding in reading						
Stimulus locked	Moderate frequency	280 to 375 ms	0.007	-683.6	0.001	0.002
	Low frequency	280 to 455 ms	0.011	-904.9	0.002	0.003



Neurophysiological Correlates of Fast Mapping of Novel Words in the Adult Brain

Marina J. Vasilyeva^{1,2*}, Veronika M. Knyazeva^{1,2}, Aleksander A. Aleksandrov^{1,2} and Yury Shtyrov^{2,3}

¹Department of Higher Nervous Activity and Psychophysiology, Saint Petersburg State University, Saint Petersburg, Russia, ²Laboratory of Behavioral Neurodynamics, Saint Petersburg State University, Saint Petersburg, Russia, ³Center of Functionally Integrative Neuroscience, Aarhus University, Aarhus, Denmark

OPEN ACCESS

Edited by:

Björn H. Schott,
Leibniz Institute for Neurobiology
(LG), Germany

Reviewed by:

Jasmin M. Kizilirmak,
German Center for
Neurodegenerative Diseases (DZNE),
Germany
Raphael Fargier,
Aix-Marseille Université, France

*Correspondence:

Marina J. Vasilyeva
marinajv@list.ru

Specialty section:

This article was submitted to *Speech and Language*, a section of the journal *Frontiers in Human Neuroscience*

Received: 18 March 2019

Accepted: 15 August 2019

Published: 19 September 2019

Citation:

Vasilyeva MJ, Knyazeva VM, Aleksandrov AA and Shtyrov Y (2019) Neurophysiological Correlates of Fast Mapping of Novel Words in the Adult Brain. *Front. Hum. Neurosci.* 13:304. doi: 10.3389/fnhum.2019.00304

Word acquisition could be mediated by the neurocognitive mechanism known as fast mapping (FM). It refers to a process of incidental exclusion-based learning and is believed to be a critical mechanism for the rapid build-up of lexicon, although its neural mechanisms are still poorly understood. To investigate the neural bases of this key learning skill, we used event-related potentials (ERPs) and employed an audio-visual paradigm that included a counterbalanced set of familiar and novel spoken word forms presented, in a single exposure, in conjunction with novel and familiar images. To define learning-related brain dynamics, passive auditory ERPs, known to index long-term memory trace activation, were recorded before and after the FM task. Following the single FM learning exposure, we found a significant enhancement in neural activation elicited by the newly trained word form, which was expressed at ~200–400 ms after the word onset. No similar amplitude increase was found either for the native familiar word used as a control stimulus in the same learning paradigm or for similar control stimuli which were not subject to training. Topographic analysis suggested a left-lateral shift of the ERP scalp distribution for the novel FM word form, underpinned by fronto-temporal cortical sources, which may indicate the involvement of pre-existing neurolinguistic networks for mastering new word forms with native phonology. Overall, the near-instant changes in neural activity after a single-shot novel word training indicate that FM could promote rapid integration of newly learned items into the brain's neural lexicon, even in adulthood.

Keywords: brain, event-related potentials, language, fast mapping, word, semantic, learning, acquisition

INTRODUCTION

From birth and throughout their entire lives, human beings learn vast amounts of new and diverse information. This is especially true in the domain of language where incredibly rapid word learning processes enable efficient mother tongue acquisition during childhood as well as mastering a second language or professional lexicon later in life. Despite numerous studies, identification of distinct neurobiological indices and putative mechanisms of such rapid word learning remains challenging.

A large number of studies have posited that rapid new word acquisition could be mediated by the neurocognitive mechanism known as *fast mapping* (FM). FM refers to a process of incidental exclusion-based learning which promotes rapid integration of newly learned items into cortical memory networks (Sharon et al., 2011; Coutanche and Thompson-Schill, 2014; Atir-Sharon et al., 2015). It was first described by Carey and Bartlett (1978) in their seminal “chromium study” with young children. These authors characterized FM as a cognitive mechanism whereby, under conditions of experimentally created ambiguity, one can infer the meaning of a new word on the principle of mutual exclusivity, and the memory trace of such newly formed associations are established and maintained even after a single exposure to the novel item. The authors noted that, for a successful mapping, a child must be able to perform “referent selection” and “referent retention” corresponding to the newly learned word. In their experiment, Carey and Bartlett presented 3–4 year-old children with two trays, one of which was red and the other one olive, and asked to bring them the “chromium” tray. Children brought the olive tray, concluding that the new word “chromium” refers to this previously unknown color. In addition, one week later, children successfully chose the “chromium” tray among six different trays, demonstrating long-term retention of the representation of the newly learned word. Further studies revealed that young children are able to reproduce new words even one month after their single presentation (Markson and Bloom, 1997; Kalashnikova et al., 2014; but see Horst and Samuelson, 2008) and that with maturation the FM mechanism becomes even more efficient.

It is now assumed that FM is a critical mechanism serving the rapid build-up of lexicon, particularly at early stages of language acquisition. Since the pioneering work of Carey and Bartlett (1978), FM has been investigated quite extensively and has been described not only in humans but also in other mammals [for example, in primates (Cook and Fagot, 2009), dogs (Kaminski et al., 2004; Pilley and Reid, 2011)] and even in birds (Cook and Fagot, 2009). However, despite numerous studies, a considerable amount of contradictory results is still evident. Moreover, only few studies evaluated FM in adults, using predominantly behavioral measures. Finally, the exact neural underpinnings of this trait remain poorly understood, as neurophysiological research in this field has been limited.

Sharon et al. (2011) reported normal learning of new word forms under FM procedure in four middle-aged patients with anterograde amnesia following hippocampal damage: their memory performance did not differ from that in healthy matched controls not only after a 10 min delay, but also one week later, whereas under a standard explicit encoding (EE) condition (i.e., not inference-based direct instruction) these individuals showed impaired explicit memory at both delays. On the other hand, patients with anterior temporal lobe (ATL) lesions (but intact hippocampi) revealed no advantage for FM condition. These and other similar results suggested FM as a hippocampally-independent learning mechanism promoting rapid neocortically-based memory formation. However, other studies in patients with hippocampal injury using slightly modified paradigms failed to replicate FM benefits over the

EE procedure (Smith et al., 2014; Warren and Duff, 2014; Warren et al., 2016).

In spite of these contradictory patient studies, recent findings in healthy young adults confirmed that learning through FM may accelerate rapid integration of newly learned items into cortical memory networks (Coutanche and Thompson-Schill, 2014), with an implicit memory measure (reaction time in a lexical task, applied after a 10-min delay as well as on the following day) revealing strong lexical competition following the FM learning procedure, while no similar evidence of lexical integration was found for EE condition. The authors proposed this pattern—rapid lexical integration of newly learned items manifest to a greater extent after incidental learning—to be a behavioral signature of FM (Coutanche and Thompson-Schill, 2014; Coutanche and Koch, 2017).

Several neuroimaging studies have claimed that FM may be linked to distinct neuroanatomical substrates. In one study the retrieval of semantic associations acquired through FM and EE conditions by healthy young subjects was measured during four alternative forced-choice recognition test using BOLD-fMRI (Merhav et al., 2015). Results indicated a specifically increased activity in ATL during retrieval in FM. Moreover, whereas a typical overnight strengthening of vmPFC engagement and vmPFC-hippocampal-neocortical interactions were apparent for EE, reflecting slower-rate consolidation processes, no similar increase was found for FM learning. The authors concluded that associative semantic learning through FM could be supported by the ATL as a critical hub enabling direct neocortical learning, bypassing the hippocampus and consolidation stage. This and other studies (e.g., Atir-Sharon et al., 2015) emphasize FM as a key neurocognitive mechanism enabling rapid neocortical plasticity to create novel semantic representations, largely supported by the temporal lobe without or minimal hippocampal involvement.

Such studies, therefore, propose a hippocampally-independent route of rapid cortical learning, without the slower consolidation stage. This somewhat contradicts the mainstream memory theories postulating that the initial fast stages of learning are hippocampus based while the neocortical memory systems are slower and require at least an overnight consolidation stage to form new representations (McClelland et al., 1995). In their seminal work, Davis and Gaskell (2009) applied principles from the Complementary Learning Systems (CLS; McClelland et al., 1995; McClelland, 2013) model of memory to brain mechanisms of word learning. According to this complementary systems account of word learning, there are two stages of lexical acquisition: (i) rapid initial encoding, largely supported by medial temporal lobe (MTL) and hippocampus, that is followed by (ii) slow lexical consolidation achieved offline in neocortex. Generally, this framework suggests that complementary systems in the hippocampus and neocortex maintain the ability to acquire new words and integrate them with existing linguistic knowledge for further retention. This framework is able to provide detailed explanations about the role of long-term memory processes in word learning; on the other hand, it pays little attention to ultra-rapid initial stages of new word forms acquisition, at least some of which have been

suggested to bypass the two-stage route and to be instantiated in the neocortex directly. While not refuting the CSL account as such, these studies suggest that it needs a certain revision to account for the rapid neocortical memory-trace formation under FM conditions.

Notably, the vast majority of studies evaluating FM used predominantly behavioral or sluggish hemodynamic measures. While fMRI is an excellent spatially-precise neuroimaging tool, its low temporal resolution makes it less optimal to study rapidly changing neural dynamics underlying language processing and fast plastic changes in brain circuits. Electroencephalography (EEG), on the contrary, can track neuronal electric activity with high temporal precision, which allows scrutinizing neural processes of language learning and comprehension on a millisecond scale. To find electrophysiological correlates of word learning, the most commonly used event-related potential (ERP) component has been the N400, a negative-going event-related brain response linked to lexical and semantic features of verbal stimuli (Kutas and Federmeier, 2011). For instance, N400 dynamics was assessed in healthy adults as they performed a contextual word-learning task in which they were required to derive the meaning of a novel word from a linguistic context provided by a few sentences (Mestres-Missé et al., 2007). After a few exposures to novel words, the N400 amplitude (initially elevated, as is typically the case for nonsense words) became statistically indistinguishable from that elicited by familiar real words, suggesting a rapid neural acquisition effect for the novel items.

Some developmental N400 studies have shown that the ability to quickly map new words and retain those representations develops quite early in childhood (Friedrich and Friederici, 2008). Children as young as 6 months old could quickly associate a new word form with the corresponding referent demonstrated in a picture—word mismatch N400 paradigm (Friedrich and Friederici, 2011). A similar study with 20-month-olds indicated substantial differences in FM efficiency in relation to child's productive vocabulary size during the period of vocabulary spurt (von Koss Torkildsen et al., 2008): children with high productive vocabularies displayed a significant N400 incongruity effect for violations in word-object mappings.

Importantly, the N400 likely reflects not the process of memory trace activation or word learning *per se*, but rather the integration of the stimulus items (such as old or new lexicon elements) in the broader context of a sentence (Friederici, 2002). On the other hand, earlier ERP components, elicited by single words outside any context in passive auditory exposure, have been shown to directly reflect lexical and lexico-semantic access processes starting from 50 to 150 ms after the auditory information allows for word identification (Shtyrov et al., 2005, 2010; Shtyrov and Pulvermüller, 2007). Several ERP studies in adults using passive perceptual learning paradigms revealed a significant increase of early electrophysiological activity in fronto-temporal cortical networks, indexing rapid learning of novel word forms after a mass exposure (Shtyrov, 2011; Kimppa et al., 2015). Furthermore, the magnitude of this brain response increase for novel word forms was predictive of further recall and recognition of the newly acquired items, supporting the notion

that such enhanced neural activity is a genuine neural correlate of the learning process (Kimppa et al., 2015). Similar findings were demonstrated in school-age children using MEG (Partanen et al., 2017), whereas this neurophysiological pattern of rapid memory trace build-up could not be found in children with dyslexia (Kimppa et al., 2018).

While such studies have suggested rapid plastic changes during word acquisition, along with neurophysiological indices of new memory trace build-up, they predominantly did not address the FM mechanism as a single-shot exclusion-based learning. They often used a series of paired word-picture presentations, story-like sentential context, direct explicit instruction or mass repetition. Overall, the neurophysiological underpinnings (and electrophysiological correlates in particular) of FM as a special form of learning still remain largely unexplored. Addressing them was the goal of the present study.

To study the neural correlates of this early implicit learning mechanism in the adult brain, we designed an experimental procedure that could model the process of rapid new word acquisition in a naturalistic FM setting. As an incidental, exclusion-based learning, FM implies inferring a word's meaning from the existing semantic context *via* "disjunctive syllogism" cognitive process (Halberda, 2006; Coutanche and Thompson-Schill, 2014; Atir-Sharon et al., 2015). Thus, we implemented an audio-visual FM learning paradigm that included counterbalanced combinations of familiar and novel words presented auditorily in pseudorandom order in conjunction with novel and familiar images. Similar to the conventional behavioral FM studies, the subject was asked to choose a new object defined by a previously unfamiliar word form, which could only be achieved by excluding other, familiar stimuli. In contrast with the vast majority of previous neuroimaging studies, only a single trial was allowed to carry out this task. This picture-word paradigm was combined with short passive EEG recording sessions run before and after the FM task since passive ERPs are known to reflect memory-trace activation and build-up (Shtyrov, 2012). We hypothesized that rapid formation of word-object associations *via* FM would be indicated by enhanced ERPs dynamics as a result of training exposure. As control conditions, we used, on the one hand, an acoustically similar familiar word undergoing similar single-shot selection task, and, on the other hand, other items that were not subject to FM training.

MATERIALS AND METHODS

Participants

Twelve monolingual native Russian speakers participated in the study [mean age = 23 ($SD = 3.9$); range 18–30 years; five men]. All were right-handed (Edinburgh inventory; Oldfield, 1971) with normal or corrected to normal vision and no record of neurological diseases. All participants were informed about the experimental procedure and signed a consent form. The study was approved by the Ethics Committee of Saint Petersburg State University and conducted in accordance with the Helsinki Declaration.

Stimuli

Acoustic Stimuli

Four acoustically and phonetically similar consonant-vowel-consonant (CVC) triphones were used as stimuli: two of them were meaningful Russian words ($[kʲit]$ —whale and $[kot]$ —cat) and the other two were phonologically legal novel word-forms ($[kʲet]$ and $[kat]$). The stimuli were recorded using a monolingual native Russian female speaker and processed in Adobe Audition 3.0 (Adobe Systems Inc., San Jose, CA, USA) and Praat v.6.0.40 (Boersma, 2002) software. Acoustic properties of the stimuli (duration, intensity, F0) were maximally matched, all stimuli were 413 ms in duration.

Visual Stimuli

Visual stimuli consisted of two-dimensional pictures of 11 familiar animals and one unknown (unreal) creature taken from the Microsoft Clipart Collection. The mean angular size of the pictures was equal to 3.5° . Acoustic and visual stimuli were presented using the Psytask software (v. 1.41.2; Mitsar Ltd, St. Petersburg, Russia) running on a Windows computer.

Experimental Design and Procedures

Experimental Design

Experimental design included an *FM exposure* and two *passive sessions* that were run immediately before and after the FM. Experiment started with a short *practice session* aimed to familiarize subjects with the task, using other stimuli than those in the main task.

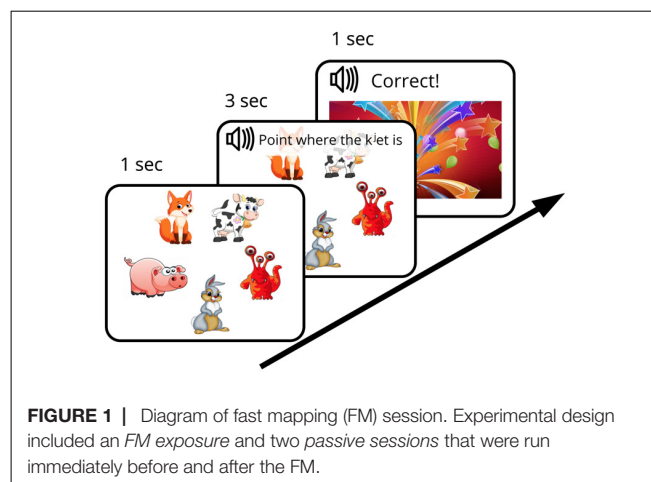
Passive Session

Passive sessions were run twice: before and after the FM procedure (which will be described below). During both passive sessions, EEG was continuously recorded. The subjects were seated in a chair facing the computer monitor located 1 m in front of them. They were instructed to pay no attention to the sound stimuli and to watch a silent video film. The acoustic stimuli (two meaningful words and two novel word-forms) were binaurally presented through the headphones at 60 dB SPL. Each stimulus was presented 25 times in pseudorandom order such that the same stimulus was not repeated twice in a row. The stimulus onset asynchrony (SOA) was jittered randomly between 1,000 and 1,100 ms.

To investigate the ERPs dynamics of the word-object association formation, we compared ERPs obtained in response to the novel words between the two passive sessions, i.e., before and after the FM condition. As a control condition, we used, on the one hand, the ERPs obtained in response to the acoustically similar previously familiar word undergoing the similar single-shot selection task, and, on the other hand, two other items (familiar word and unfamiliar pseudo-word) that were not subject to FM training, but were used as a control for the mere repetition of stimuli in the passive session.

Fast Mapping

We designed an FM procedure with the aim of mimicking the earlier behavioral investigations (such as the original “chromium” study described above) as closely as possible



while adapting them to an EEG experiment setting with well-matched control stimuli and strictly defined recognition points allowing for precise ERP time-locking. The designed FM procedure was aimed to investigate the subject’s ability (under the conditions of experimentally created ambiguity) to infer, by exclusion, the referent of a novel word from a brief single exposure and to store this newly formed word-object mapping in memory for later use. To this end, an audio-visual FM paradigm with preferential pointing task was applied (Spiegel and Halberda, 2011). The subject was asked to identify one unknown object among pictures of familiar ones, all presented simultaneously on the screen. First, five objects, arranged in a circle and counterbalanced for position, appeared on the screen on a white background. After a short delay (~ 1 s) an auditory request (3 s long) relating to one of the objects was made (“Point, where X is”) and the subject had to identify which object was being referred to (in case of the novel word, this was only possible by excluding other, familiar, objects) and to point to it. The FM session started with a short *practice session* that included two trials with familiar word-object combinations [e.g., $[gusʲ]$ (goose), $[konʲ]$ (horse)], which were not used in the experiment proper. After that, one trial with familiar word $[kot]$ (cat) paired with a known visual item and one FM trial with the target new word form ($[kʲet]$) paired with an image of the novel visual item, displayed together with four familiar objects, were presented (**Figure 1**). After the subjects succeeded in referent selection (familiar or target new word), they were greeted by the experimenter and a colorful picture of firework appeared on the screen (for 1 s). Whereas one real word and one novel word form from the set of four underwent this FM procedure, the other two items (acoustically similar real word and pseudo word) were used as control stimuli, i.e., they were present in the passive ERP recordings but not in the FM condition. The small number of stimuli was used in order to both approximate the early behavioral designs, which typically used a single novel item in an exposure session, and to avoid any potential interference that could arise should multiple items be used.

Data Analysis

EEG was continuously recorded using a 32-channel Mitsar EEG set-up and WinEEG software (Mitsar Ltd) with 500 Hz sampling rate and pass-band of 0.01–150 Hz. Ag/AgCl electrodes were mounted in an extended 10-20-system electrode cap; hardware-linked earlobe electrodes were used as the reference channel. To control for eye-movements, horizontal and vertical electro-oculograms (EOG) were recorded. The impedance of the electrodes did not exceed 10 kOhm. The signal was bandpass filtered between 0.5 and 45 Hz offline.

EEG data obtained during the passive sessions were epoched from 100 ms before to 700 ms after the stimulus onset. The baseline was corrected using a 100-ms pre-stimulus interval. EEG epochs in which the EEG or EOG signal amplitude exceeded $\pm 100 \mu\text{V}$ on any of the electrodes were omitted. The average number of trials remaining after artifact removal was $21.4 \pm \text{SD } 2.5$ out of the total of 25 per type. Two subjects were excluded from the final dataset due to excessive artifacts in the EEG recordings; thus, 10 subjects were included in statistical analysis.

Amplitude analysis was carried out for the fronto-central electrode cluster where the auditory evoked responses are typically maximal: F3, Fz, F4, FC3, FCz, FC4, C3, Cz, and C4. Visual data inspection revealed the presence of several pronounced peaks in a broad 100–500 ms time window. Since we were agnostic with respect to the latency when FM effects might occur, we opted for an unbiased data-driven approach and split the epochs into equal 100-ms bins, and performed an exploratory analysis of each of the four bins. Mean ERP amplitudes were calculated over 100 ms time intervals, starting from 100 ms, when the stimuli could be differentiated acoustically. These were analyzed for each stimulus type separately using the repeated measures analysis of variance (rmANOVA; SPSS v. 21, IBM Corporation, New York, NY, USA) with *Session* (before/after FM), *Electrode* (frontal, central-frontal, central) and *Location* (left, right, medial) factors. Greenhouse-Geisser correction was applied whenever the sphericity assumption was violated; multiple comparisons were corrected for using Bonferroni corrections where necessary. Average ERPs for each stimulus type were calculated by combining epochs of each familiar or novel word separately. Effect sizes were calculated using

partial eta squared (η_p^2 ; SPSS v. 21, IBM Corporation, New York, NY, USA).

Low-resolution electromagnetic tomography (LORETA; Pascual-Marqui et al., 1994) images were obtained by estimating the current source density distribution of brain electric activity on a dense grid of 2,394 voxels at 7-mm spatial resolution applied to the digitized Talairach human atlas (Talairach and Tournoux, 1988). To this end, the group-average difference between the ERPs recorded before and after FM session was submitted to LORETA. Group-average data were used since they benefit from a much-increased signal-to-noise ratio that source analysis algorithms are highly sensitive to, which, in turn, could somewhat compensate for the low resolution of the EEG technique applied.

RESULTS

The FM session was completed successfully by all subjects. Here, we present the results of comparing the ERP data collected in passive sessions run before and after the FM learning condition. ERPs were recorded to passively presented auditory stimuli, including familiar and novel items used in the FM session and untrained control stimuli.

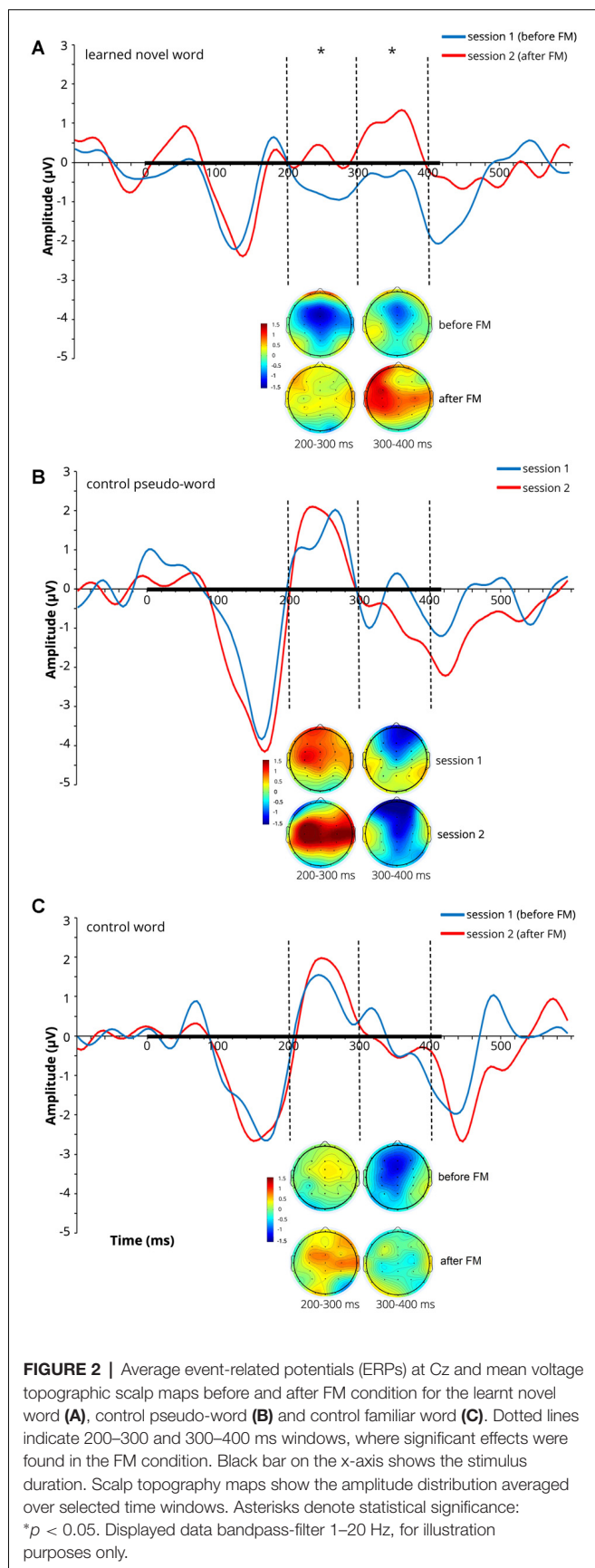
We split the epochs into equal 100-ms bins during the time when most typical word-related ERPs might take place and performed an exploratory analysis of each of the 4 bins. Analysis of ERP data using rmANOVA is presented in **Table 1**. Analysis of data from the 100–200 ms window (which showed a negative peak in the subtraction curve with 155 ms average latency) revealed no significant main effects of *Session*, *Electrode* or *Location* as well as no interaction effects.

At later latencies, a significant main effect of *Session* on the ERP amplitude was found over both the 200–300 ms ($F_{(1,9)} = 5.398$, $p = 0.045$, $\eta_p^2 = 0.375$) and 300–400 ms ($F_{(1,9)} = 8.428$, $p = 0.018$, $\eta_p^2 = 0.484$) windows for the learnt novel word form reflecting an amplitude increase after the FM condition (note that both windows showed positive-going peaks in the subtraction curves at 250 ms and 360 ms). No significant main effects of *Electrode* and *Location* and no interaction effects were found in those time windows. Average ERPs at Cz and mean voltage topographic scalp maps before and

TABLE 1 | Analysis of variance (ANOVA) results.

SM	100–200 ms	200–300 ms	300–400 ms	400–500 ms
Session	$F_{(1,9)} = 0.545$ $p = 0.479$ $\eta_p^2 = 0.057$	$F_{(1,9)} = 5.398$ $p = 0.045^*$ $\eta_p^2 = 0.375$	$F_{(1,9)} = 8.428$ $p = 0.018^*$ $\eta_p^2 = 0.484$	$F_{(1,9)} = 0.995$ $p = 0.345$ $\eta_p^2 = 0.100$
Electrode	$F_{(1,319,11,875)} = 1.542$ $p = 0.247$ $\eta_p^2 = 0.146$	$F_{(1,302,11,716)} = 0.189$ $p = 0.736$ $\eta_p^2 = 0.021$	$F_{(1,122,10,099)} = 3.167$ $p = 0.103$ $\eta_p^2 = 0.260$	$F_{(1,129,10,162)} = 2.351$ $p = 0.155$ $\eta_p^2 = 0.207$
Location	$F_{(1,781,16,025)} = 0.320$ $p = 0.706$ $\eta_p^2 = 0.034$	$F_{(1,736,15,627)} = 0.434$ $p = 0.628$ $\eta_p^2 = 0.046$	$F_{(1,691,15,222)} = 3.407$ $p = 0.066$ $\eta_p^2 = 0.275$	$F_{(1,756,15,800)} = 7.827$ $p = 0.005^*$ $\eta_p^2 = 0.465$

Analysis included data from nine electrodes, with factors of *Session*, *Electrode* and *Location*. No significant main effects of interaction were revealed over all selected time windows. Significant results ($p < 0.05$) are highlighted by an asterisk (*).



after FM condition for the target novel word form are shown in **Figure 2A**.

A significant effect of *Location* ($F_{(1,756,15,800)} = 7.827$, $p = 0.005$, $\eta_p^2 = 0.465$) was found over the 400–500 ms window. Multiple pair-wise comparisons revealed the significant amplitude enhancement for the novel and familiar word forms in the left hemisphere as compared to the right hemisphere ($p = 0.004$). No significant main effects of *Session* and *Electrode* and no interaction effects were found in this time interval.

Difference wave (at Cz) obtained by subtracting the ERPs for the learnt novel word used in passive session 1 (before FM) from those in passive session 2 (after FM) and corresponding difference topographic scalp maps are presented in **Figure 3**. Additional figures for control pseudoword and control familiar word are presented in **Supplementary Figure S1**.

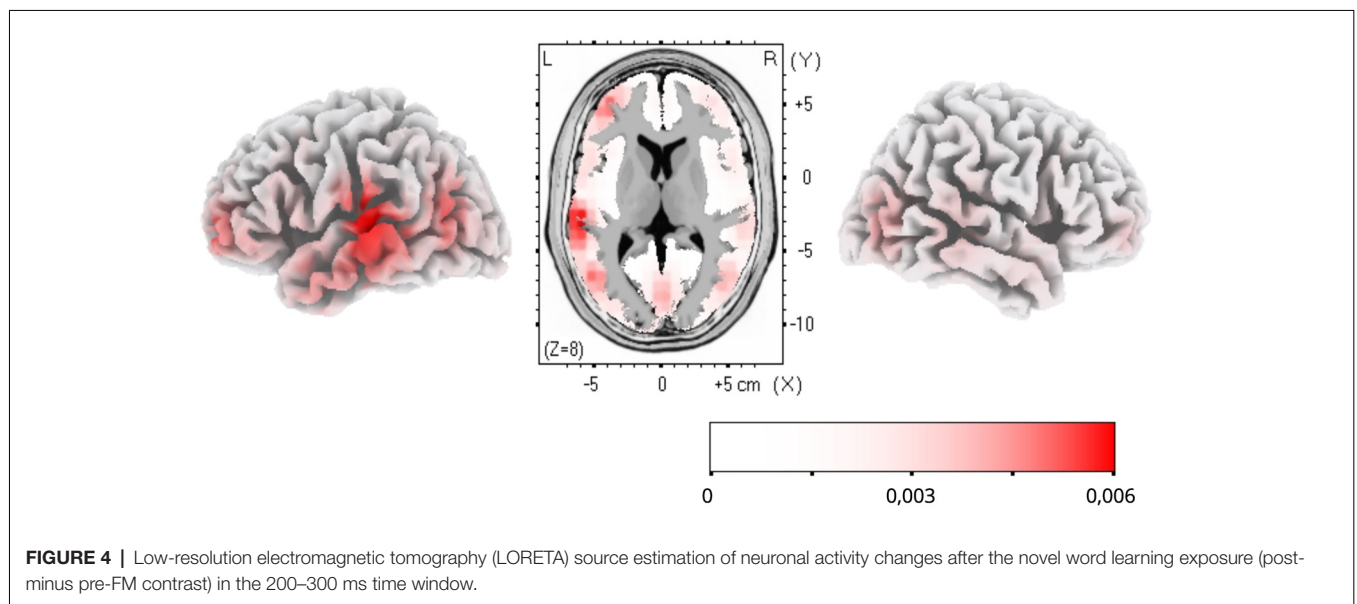
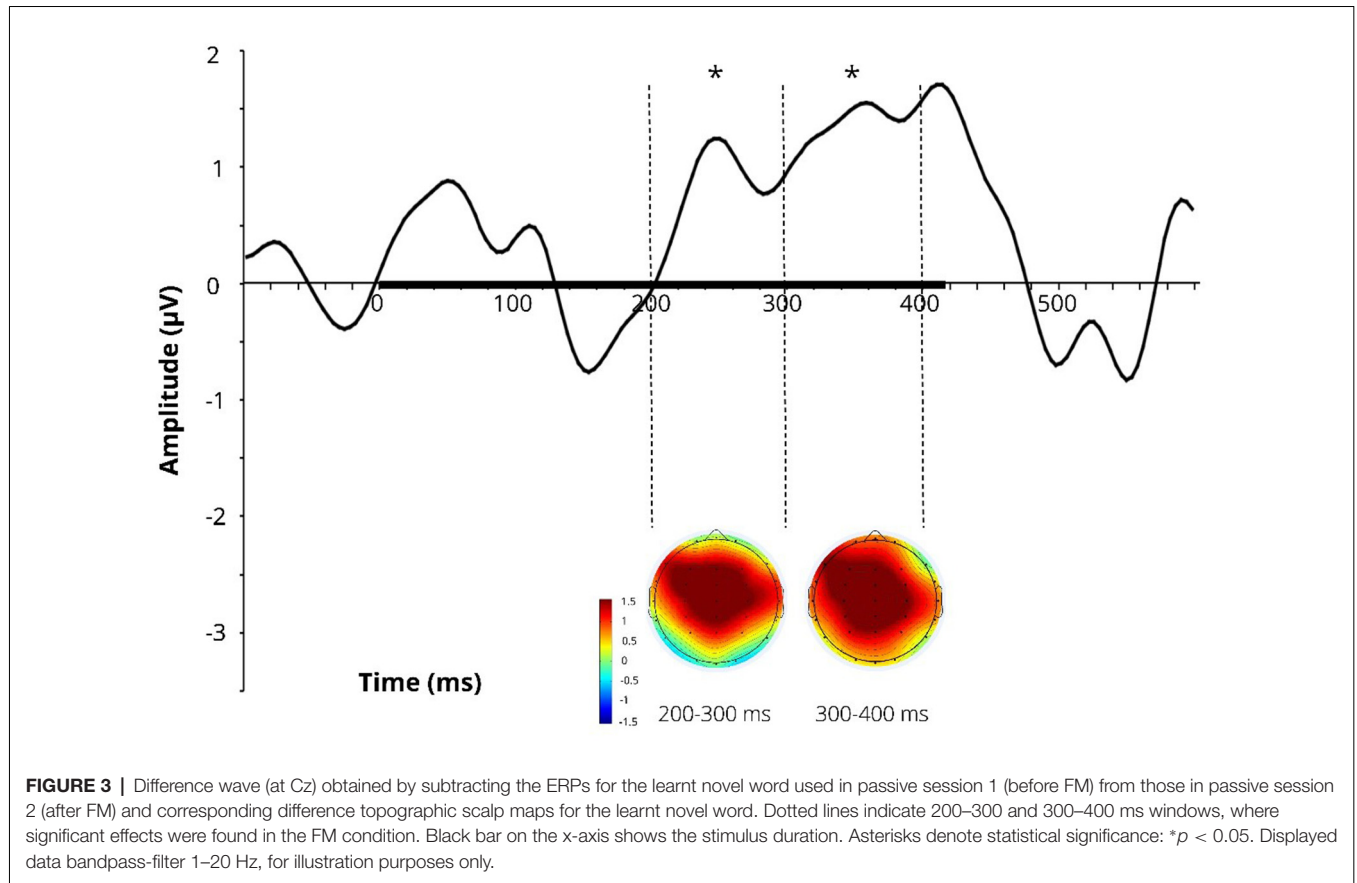
In addition, the analysis of N400 component indicated that the N400 peaked at 416 ± 36 ms in the before-FM condition. Thus, we conducted an *ad hoc* analysis of N400 component in the a 100-ms window centered on this peak (i.e., 366–466 ms), which confirmed the significant *Location* main effect ($F_{(1,878,16,898)} = 10.116$, $p = 0.001$, $\eta_p^2 = 0.529$) that was obtained at 400–500 ms time bin. However, no *Session* effect was observed ($F_{(1,9)} = 3.938$, $p = 0.078$, $\eta_p^2 = 0.304$).

To estimate cortical sources of the training-related ERP dynamics, LORETA computation in Talairach space was applied to group-average subtractions of ERP traces before and after FM condition in the time windows of significant ERP effects. The LORETA results are shown in **Figure 4**. Maximal activity was observed in the left temporal cortex (peaking in BA21), with a less pronounced source in the left anterior prefrontal cortex. No differences were found for the control familiar word used in the FM condition or control items given in passive sessions only (**Figures 2B,C**).

DISCUSSION

This study aimed at delineating neural correlates of FM of phonologically and semantically novel words through a single-shot exposure in a naturalistic inference-based learning scenario. We found a significant enhancement in ERP amplitudes elicited by a native novel word form following this simple semantic learning task. This enhancement was found using passive auditory ERPs, known to be an index of automatic memory trace activation (Shtyrov et al., 2005, 2010; Shtyrov and Pulvermüller, 2007), and was maximal over 200–400 ms after the word onset, i.e., shortly after the words could be identified as distinct and even before their offset. Notably, no difference was found for either the native familiar words used in the same experimental conditions or for the control phonologically legal pseudoword given in passive sessions only. These different types of control conditions rule out a possibility that the current ERP dynamics could simply be explained based on physical stimulus repetition; instead, the observed change in the brain’s response seems to be best interpreted as a specific consequence of the FM procedure.

Previous studies have shown that formation of neural memory traces for novel spoken word forms with native



phonology could be captured after a mass exposure, with multiple (sometimes dozens or hundreds) repetitions (Shtyrov, 2011; Kimppa et al., 2015; Partanen et al., 2017) whereas the current rapid ERP dynamics was revealed after a one-trial exposure to new word-picture pairs. Given the control

conditions employed (involving FM word as well as non-FM word and pseudoword presented in the passive session only), the present result points toward semantic context advantage in the rapid formation of novel memory traces for words.

These results are similar to some previous investigations that found online changes of brain dynamics in the process of novel meaning acquisition implemented through word-picture associations or sentential context (e.g., Breitenstein et al., 2005; Mestres-Missé et al., 2007). Still, to our knowledge, the current data is the first electrophysiological *evidence of FM proper, as the process of exclusion-based inference learning implemented in a single shot*. While previous studies focused on using purely behavioral or slow hemodynamic measures, here, this neurophysiological signature of the processes underlying rapid word acquisition in healthy adults is documented as a dynamic enhancement of electrophysiological response. This enhancement is most likely underpinned by an automatic activation of the newly created word memory trace, realized as a robust neuronal circuit formed in the process of associative learning (Pulvermüller et al., 2001; Aleksandrov et al., 2011).

Some previous studies (Wilding, 1999, 2000) recorded ERPs during recognition memory tasks aimed to differentiate old (studied) and new (untrained) visually presented words. ERPs to words judged correctly to be old were more positive (at left parietal sites) than new ones. This left-lateralized old/new effect is known to index the process of recollection (retrieval) from episodic memory. Moreover, its magnitude appeared to be related to the amount/quality of information retrieved from memory. Our findings are somewhat similar in the amplitude patterns, even though our study was designed with a different paradigm and assessment technique: in conjunction with FM paradigm (not aimed at memorizing as such, but rather at incidental learning through inference), we used passive auditory ERPs that are known to be a neurophysiological index of automatic memory trace activation and build-up (Shtyrov, 2012; Kimppa et al., 2015; Partanen et al., 2017) rather than active retrieval. Notably, the magnitude of this brain response increase for novel word forms is predictive of further recall and recognition of the newly acquired items (supporting the notion that such enhanced neural activity is a genuine neural correlate of the learning process, Kimppa et al., 2015), which again bears clear similarity to the Wilding's studies above.

The topographic analysis of the amplitude distribution of these ERP changes suggested a more pronounced left laterality effect of FM on novel word-form learning. Additionally, analysis of cortical activity sources using LORETA confirmed that the learning dynamics was underpinned by sources in the left temporal and inferior-frontal cortices, indicating that this response enhancement is likely underpinned by the perisylvian neural network specialized in native language processing. Overall, it may be proposed that FM may induce rapid neocortical plasticity in healthy adult brain by engaging pre-existing language neural networks for mastering new word forms with native phonology (Shtyrov, 2011; Kimppa et al., 2015; Partanen et al., 2017). On a more cautious note, since, for increased SNR which source-analysis algorithms are highly sensitive to, grand-average data were used for LORETA estimations, no statistical verification of the source activation is possible and the results should, therefore, be treated as indicative of an average "center of gravity" of cortical generators, rather than definitive. That implies that the present source

analysis outcomes should be treated with extreme caution and must be verified in future studies (using, e.g., combined EEG/MEG with MR-based cortical models) with respect to the exact effect origins. That said, the left temporo-frontal distribution of the LORETA effects found is well compatible with existing knowledge of the cortical language and learning systems and thus still provides a useful illustration of putative neural underpinnings of the FM mechanism.

ERP dynamics observed here exhibited some differences from the earlier investigations. First, in our study the earliest registered activity manifesting differential dynamics after semantic learning task was around 200–400 ms with a left-central distribution of positive polarity. Several previous studies demonstrated earlier rapid lexical effects in auditory ERPs starting from already ~50–100 ms with predominantly negative polarity deflections and less lateralized fronto-central distribution (e.g., Shtyrov et al., 2010). Those studies, however, predominantly time-locked ERPs to word-recognition points located at word offsets, whereas here the word identification became possible in the initial CV transition; while it is more difficult to precisely indicate these transition points, they likely occur within the first ~200 ms after onset, implying that, in terms of word recognition, our latencies are comparable with previous studies. Interestingly, no significant *Session* effect was observed in the N400 time window. A potential explanation for this is that the typical N400 effects reflect the integration of single words into wider (sentential) context. Here, however, only single words were presented outside any phrases, and the putative lexico-semantic activation took place at an earlier time interval rather than in the typical N400 range. As for polarity, at least one earlier ERP experiment reported an increase in frontal positivity during rapid language learning (Shtyrov, 2011), which is what we found here as well, even though we implemented a rather different learning paradigm and stimuli. Finally, previous fMRI work suggested ATL as the primary hub for implicit FM learning (e.g., Merhav et al., 2015), whereas our LORETA results suggest a more posterior-superior temporal lobe activation; this divergence, however, cannot be resolved based on the current data: on the one hand, activation in the temporal pole is known to be unreliable in both fMRI and EEG; on the other hand, the present source reconstruction results should be treated with extreme caution as they are based on low-resolution EEG data and present a group-average picture which cannot be verified statistically. Further investigations of the exact learning-related neural dynamics and of their neuroanatomical origins are clearly needed to scrutinize these processes in more detail; one way to pursue this could be to use combined MEG/EEG with individual MR-based source reconstructions techniques.

Notably, the very brief and subject-friendly novel paradigm we have developed—based on a non-demanding single-shot learning task and short passive auditory ERP recording—allows for investigations of FM processes in diverse populations, including young children, elderly subjects or different patient groups. Future studies may apply and further develop this approach to assess learning-related neural dynamics and their deficits in different conditions, populations and experimental settings.

In this experiment, we have followed the strategy of the original behavioral FM studies that typically used a single novel item in one exposure session. The logic for this is two-fold. First, it was aimed at mimicking the earlier investigations as closely as possible, while adapting them to an EEG experiment setting with two types of control stimuli and strictly defined recognition points, to allow for precise ERP time-locking. Second, to our knowledge, no previous study has electrophysiologically investigated single-shot learning in classical FM situation in its strict sense; therefore, in this first endeavor, we opted for a safe approach avoiding any potential interference from using multiple items. Previous EEG studies did not address the FM mechanism in its strict sense as a one-trial inference-based learning; instead, they used series of paired word-picture presentations of the same items, story-like sentential context with multiple item occurrences or even mass stimulus repetition. In contrast to the vast majority of such previous behavioral and neuroimaging studies, in our study, only a single trial was allowed to carry out the FM task. This strategy has obviously proved to be fruitful in the present case. Indeed, on the one hand, we found a significant enhancement in ERPs amplitudes after a one-trial exposure to the newly inferred item, which appears to be an important advance on its own. On the other hand, however, such a restricted stimulus design does not easily allow for general conclusions concerning the current findings and the neurobiological mechanisms involved. Therefore, taken at face value, the current result should still be treated with caution. We suggest that future studies should expand the approach developed here to possibly use several FM items to both validate our results and generalize them.

On a similarly cautious note, even though the effect sizes obtained here are fairly good and results clearly demonstrate significant ERP changes following the FM procedure, we suggest that, for reliability and reproducibility, future studies could use larger subject samples than the that employed here. A somewhat more difficult question relates to the number of trials employed in the present passive sequence. Considering that the massive stimulus repetition *per se*, as discussed above, leads to memory trace build-up even in passive designs without any semantic training, we limited the passive sessions to 25 trials only. This number is on the low end of scale for a reliable ERP (although this *per se* does not undermine the present result); future studies could circumvent this issue by using multiple tokens (e.g., 2–4, with 25 repetitions each) that can be combined to produce ERPs with higher SNRs.

In sum, the results of the current study suggest that the FM mechanism of word acquisition, well established in previous behavioral research, promotes incidental rapid integration of new associations into neocortical lexico-semantic networks in healthy adult brain as indicated by the rapid changes in ERPs

REFERENCES

Aleksandrov, A. A., Boricheva, D. O., Pulvermüller, F., and Shtyrov, Y. (2011). Strength of word-specific neural memory traces assessed electrophysiologically. *PLoS One* 6:e22999. doi: 10.1371/journal.pone.0022999

present after a brief single exposure to a novel item. Future studies are needed to validate the current findings and generalize them to other stimulus types, languages and experimental groups, to clarify the neuroanatomical underpinnings of this mechanism as well as to scrutinize these neural FM processes in typical and atypical development.

DATA AVAILABILITY

The raw data supporting the conclusions of this manuscript will be made available by the authors, without undue reservation, to any qualified researcher.

ETHICS STATEMENT

The study was approved by the Ethics Committee of Saint Petersburg State University and was conducted in accordance with the Declaration of Helsinki. All subjects gave written informed consent.

AUTHOR CONTRIBUTIONS

AA and YS designed the study. MV and VK performed the experiment and data analysis. YS supervised this work. All authors discussed the results and contributed to the final manuscript.

FUNDING

This work was supported by the Government of Russian Federation, grant contract № 14.W03.31.0010.

ACKNOWLEDGMENTS

We would like to thank Kristina Memetova for her help in recording the stimuli used in the experiment.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnhum.2019.00304/full#supplementary-material>

FIGURE S1 | Difference wave (at Cz) obtained by subtracting the ERPs for the learnt novel word used in passive session 1 (before FM) from those in passive session 2 (after FM) and corresponding difference topographic scalp maps for the learnt novel word (A), control pseudo-word (B) and control familiar word (C). Dotted lines indicate 200–300 and 300–400 ms windows, where significant effects were found in the FM condition. Black bar on the x-axis shows the stimulus duration. Asterisks denote statistical significance: * $p < 0.05$. Displayed data bandpass-filtered at 1–20 Hz, for illustration purposes only.

Atir-Sharon, T., Gilboa, A., Hazan, H., Koilis, E., and Manevitz, L. M. (2015). Decoding the formation of new semantics: MVPA investigation of rapid neocortical plasticity during associative encoding through fast mapping. *Neural Plast.* 2015:804385. doi: 10.1155/2015/804385

Boersma, P. (2002). Praat, a system for doing phonetics by computer. *Glott Int.* 5, 341–345.

- Breitenstein, C., Jansen, A., Deppe, M., Foerster, A. F., Sommer, J., Wolbers, T., et al. (2005). Hippocampus activity differentiates good from poor learners of a novel lexicon. *Neuroimage* 25, 958–968. doi: 10.1016/j.neuroimage.2004.12.019
- Carey, S., and Bartlett, E. (1978). “Acquiring a single new word,” in *Proceedings of the Stanford Child Language Conference*, (Republished in *Papers and Reports on Child Language Development* 15, 17–29).
- Cook, R., and Fagot, J. (2009). First trial rewards promote 1-trial learning and prolonged memory in pigeon and baboon. *Proc. Natl. Acad. Sci. U S A* 106, 9530–9533. doi: 10.1073/pnas.0903378106
- Coutanche, M. N., and Koch, G. E. (2017). Variation across individuals and items determine learning outcomes from fast mapping. *Neuropsychologia* 106, 187–193. doi: 10.1016/j.neuropsychologia.2017.09.029
- Coutanche, M. N., and Thompson-Schill, S. L. (2014). Fast mapping rapidly integrates information into existing memory networks. *Exp. Psychol. Gen.* 143, 2296–2303. doi: 10.1037/xge0000020
- Davis, M. H., and Gaskell, M. G. (2009). A complementary systems account of word learning: neural and behavioural evidence. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 3773–3800. doi: 10.1098/rstb.2009.0111
- Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends Cogn. Sci.* 6, 78–84. doi: 10.1016/s1364-6613(00)01839-8
- Friedrich, M., and Friederici, A. D. (2008). Neurophysiological correlates of online word learning in 14-month olds infants. *Neuroreport* 19, 1757–1761. doi: 10.1097/wnr.0b013e328318f014
- Friedrich, M., and Friederici, A. D. (2011). Word learning in 6-month-olds: fast encoding-weak retention. *J. Cogn. Neurosci.* 23, 3228–3240. doi: 10.1162/jocn_a_00002
- Halberda, J. (2006). Is this a dax which I see before me? Use of the logical argument disjunctive syllogism supports word-learning in children and adults. *Cogn. Psychol.* 53, 310–344. doi: 10.1016/j.cogpsych.2006.04.003
- Horst, J. S., and Samuelson, L. K. (2008). Fast mapping but poor retention by 24-month-old infants. *Infancy* 13, 128–157. doi: 10.1080/15250000701795598
- Kalashnikova, M., Mattock, K., and Monaghan, P. (2014). Disambiguation of novel labels and referential facts: a developmental perspective. *First Lang.* 34, 125–135. doi: 10.1177/0142723714525946
- Kaminski, J., Call, J., and Fischer, J. (2004). Word learning in a domestic dog: evidence for “fast mapping”. *Science* 304, 1682–1683. doi: 10.1126/science.1097859
- Kimppa, K., Kujala, T., Leminen, A., Vainio, M., and Shtyrov, Y. (2015). Rapid and automatic speech-specific learning mechanism in human neocortex. *Neuroimage* 118, 282–291. doi: 10.1016/j.neuroimage.2015.05.098
- Kimppa, K., Shtyrov, Y., Partanen, E., and Kujala, T. (2018). Impaired neural mechanism for online novel word acquisition in dyslexic children. *Sci. Rep.* 8:12779. doi: 10.1038/s41598-018-31211-0
- Kutas, M., and Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Ann. Rev. Psychol.* 62, 621–647. doi: 10.1146/annurev.psych.093008.131123
- Markson, L., and Bloom, P. (1997). Evidence against a dedicated system for word learning in children. *Nature* 385, 813–815. doi: 10.1038/385813a0
- McClelland, J. L. (2013). Incorporating rapid neocortical learning of new schema-consistent information into complementary learning systems theory. *J. Exp. Psychol. Gen.* 142, 1190–1210. doi: 10.1037/a0033812
- McClelland, J. L., McNaughton, B. L., and O’Reilly, R. C. (1995). Why there are complementary learning-systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychol. Rev.* 102, 419–457. doi: 10.1037/0033-295X.102.3.419
- Merhav, M., Karni, A., and Gilboa, A. (2015). Not all declarative memories are created equal: fast mapping as a direct route to cortical declarative representations. *Neuroimage* 117, 80–92. doi: 10.1016/j.neuroimage.2015.05.027
- Mestres-Missé, A., Rodriguez-Fornells, A., and Münte, T. F. (2007). Watching the brain during meaning acquisition. *Cereb. Cortex* 17, 1858–1866. doi: 10.1093/cercor/bhl094
- 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
- Partanen, E., Leminen, A., Paoli, S., Bundgaard, A., Kingo, O. S., Krøjgaard, P., et al. (2017). Flexible, rapid and automatic neocortical word form acquisition mechanism in children as revealed by neuromagnetic brain response dynamics. *Neuroimage* 155, 450–459. doi: 10.1016/j.neuroimage.2017.03.066
- Pascual-Marqui, R. D., Michel, C. M., and Lehmann, D. (1994). Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain. *Int. J. Psychophysiol.* 18, 49–65. doi: 10.1016/0167-8760(84)90014-x
- Pilley, J. W., and Reid, A. K. (2011). Border collie comprehends object names as verbal referents. *Behav. Processes* 86, 184–195. doi: 10.1016/j.beproc.2010.11.007
- Pulvermüller, F., Kujala, T., Shtyrov, Y., Simola, J., Tiitinen, H., Alku, P., et al. (2001). Memory traces for words as revealed by the mismatch negativity. *Neuroimage* 14, 607–616. doi: 10.1006/nimg.2001.0864
- Sharon, T., Moscovitch, M., and Gilboa, A. (2011). Rapid neocortical acquisition of long-term arbitrary associations independent of the hippocampus. *Proc. Natl. Acad. Sci. U S A* 108, 1146–1151. doi: 10.1073/pnas.1005238108
- Shtyrov, Y. (2011). Fast mapping of novel word forms traced neurophysiologically. *Front. Psychol.* 2:340. doi: 10.3389/fpsyg.2011.00340
- Shtyrov, Y. (2012). Neural bases of rapid word learning. *Neuroscientist* 18, 312–319. doi: 10.1177/1073858411420299
- Shtyrov, Y., Nikulin, V., and Pulvermüller, F. (2010). Rapid cortical plasticity underlying novel word learning. *J. Neurosci.* 30, 16864–16867. doi: 10.1523/jneurosci.1376-10.2010
- Shtyrov, Y., Pihko, E., and Pulvermüller, F. (2005). Determinants of dominance: is language laterality explained by physical or linguistic features of speech? *Neuroimage* 27, 37–47. doi: 10.1016/j.neuroimage.2005.02.003
- Shtyrov, Y., and Pulvermüller, F. (2007). Early MEG activation dynamics in the left temporal and inferior frontal cortex reflect semantic context integration. *J. Cogn. Neurosci.* 19, 1633–1642. doi: 10.1162/jocn.2007.19.10.1633
- Smith, C. N., Urgolites, Z. J., Hopkins, R. O., and Squire, L. R. (2014). Comparison of explicit and incidental learning strategies in memory-impaired patients. *Proc. Natl. Acad. Sci. U S A* 111, 475–479. doi: 10.1073/pnas.1322263111
- Spiegel, C. J., and Halberda, J. (2011). Rapid fast-mapping abilities in 2-year-olds. *J. Exp. Child Psychol.* 109, 132–140. doi: 10.1016/j.jecp.2010.10.013
- Talairach, J., and Tournoux, P. (1988). *Co-Planar Stereotaxic Atlas of the Human Brain*. New York, NY: Thieme, Stuttgart.
- von Koss Torkildsen, J., Svangstu, J. M., Hansen, H. F., Smith, L., Simonsen, H. G., Moen, I., et al. (2008). Productive vocabulary size predicts event-related potential correlates of fast mapping in 20-month-old. *J. Cogn. Neurosci.* 20, 1266–1282. doi: 10.1162/jocn.2008.20087
- Warren, D. E., and Duff, M. C. (2014). Not so fast: hippocampal amnesia slows word learning despite successful fast mapping. *Hippocampus* 24, 920–933. doi: 10.1002/hipo.22279
- Warren, D. E., Tranel, D., and Duff, M. C. (2016). Impaired acquisition of new words after left temporal lobectomy despite normal fast-mapping behavior. *Neuropsychologia* 80, 165–175. doi: 10.1016/j.neuropsychologia.2015.11.016
- Wilding, E. L. (1999). Separating retrieval strategies from retrieval success: an event related potential study of source memory. *Neuropsychologia* 37, 441–454. doi: 10.1016/s0028-3932(98)00100-6
- Wilding, E. L. (2000). In what way does the parietal ERP old/new effect index recollection? *Int. J. Psychophysiol.* 35, 81–87. doi: 10.1016/s0167-8760(99)00095-1

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Vasilyeva, Knyazeva, Aleksandrov and Shtyrov. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Auditory Mismatch Negativity Response in Institutionalized Children

Irina Ovchinnikova^{1,2†}, Marina A. Zhukova^{1,2†}, Anna Luchina¹, Maxim V. Petrov¹, Marina J. Vasilyeva³ and Elena L. Grigorenko^{1,2,4,5*}

¹Laboratory of Translational Sciences of Human Development, Saint-Petersburg State University, Saint-Petersburg, Russia, ²Department of Psychology, University of Houston, Houston, TX, United States, ³Department of Higher Nervous Activity and Psychophysiology, Biological Faculty, Saint-Petersburg State University, Saint-Petersburg, Russia ⁴Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, United States, ⁵Child Study Center and Haskins Laboratories, Yale University, New Haven, CT, United States

OPEN ACCESS

Edited by:

Beatriz Martín-Luengo,
National Research University Higher
School of Economics, Russia

Reviewed by:

Jarmo Hamalainen,
University of Jyväskylä, Finland
Rick A. Adams,
University College London,
United Kingdom
Linjun Zhang,
Beijing Language and Culture
University, China

Eino Partanen,
University of Helsinki, Finland

*Correspondence:

Elena L. Grigorenko
elena.grigorenko@times.uh.edu

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to Speech
and Language, a section of the
journal *Frontiers in Human
Neuroscience*

Received: 19 March 2019

Accepted: 13 August 2019

Published: 25 September 2019

Citation:

Ovchinnikova I, Zhukova MA,
Luchina A, Petrov MV, Vasilyeva MJ
and Grigorenko EL (2019) Auditory
Mismatch Negativity Response in
Institutionalized Children.
Front. Hum. Neurosci. 13:300.
doi: 10.3389/fnhum.2019.00300

The attunement of speech perception/discrimination to the properties of one's native language is a crucial step in speech and language development at early ages. Studying these processes in young children with a history of institutionalization is of great interest, as being raised in institutional care (IC) may lead to lags in language development. The sample consisted of 82 children, split into two age groups. The younger age group (<12 months) included 17 children from the IC and 17 children from the biological-family-care (BFC) group. The older group (>12 months) consisted of 23 children from the IC group, and 25 children from the BFC group. A double-oddball paradigm with three consonant-vowel syllables was used, utilizing native (Russian) and foreign (Hindi) languages. A Mismatch Negativity (MMN) component was elicited within a 125–225 ms time window in the frontal-central electrode. Findings demonstrate the absence of MMN effect in the younger age group, regardless of the living environment. Children in the older group are sensitive to native deviants and do not differentiate foreign language contrasts. No significant differences were observed between the IC and BFC groups for children older than 12 months, indicating that children in the IC have typical phonological processing. The results show that the MMN effect is not registered in Russian speaking children before the age of 12 months, regardless of their living environment. At 20 months of age, institutionally reared children show no evidence of delays in phonetic development despite a limited experience of language.

Keywords: institutionalization, psychosocial deprivation, language development, auditory discrimination, event-related potentials, mismatch negativity, MMN

INTRODUCTION

Institutional care (IC) remains a common type of placement for children raised without biological families in a number of countries, including the Russian Federation. Detrimental effects of IC have been well documented for different developmental domains, including language development. Studies show that children with an IC history, as a group, demonstrate a lack of comprehensive utterances at the age of 30 months when exposed to severe deprivation, such as in Romanian orphanages (Windsor et al., 2007), poor sentence comprehension, working memory deficits (Desmarais et al., 2012), and lower academic performance (Vorria et al., 2014) when exposed to

institutional settings of variable quality. Documented deficits in language development have been associated with the length of institutionalization (Loman et al., 2009), especially for the receptive language domain (Eigsti et al., 2011; Desmarais et al., 2012). It has been argued that observed language deficits might be caused by the alteration of neural structures in IC children due to chronic stress and psycho-social deprivation (Eigsti et al., 2011), as well as impoverished input, a limited quantity and quality of linguistic input, and disrupted child-caregiver interactions (Windsor et al., 2007). Language is learned *via* social interaction (Kuhl et al., 2003), which institutionally reared children might be deprived of due to lack of caregivers' responsiveness, low stability of the environment, and limited amount of child-directed interactions (Muhamedrahimov et al., 2005, 2014). Therefore, the lack of social interactions can result in poorer phonetic discrimination skills in children raised in institutions.

Event-related potential (ERP) studies of institutionalized children in the Russian Federation have shown that children of 30 months and above show attenuated processing of semantic incongruities manifested in the atypical N400 component, compared to peers raised in biological families (Zhukova et al., 2015). It is argued that an atypical neural response to semantic incongruity may reflect underspecified lexical representations or altered functional connectivity in children raised in IC in Russia. Data acquired from adults who were raised in institutions in the Russian Federation suggest that detrimental effects of institutionalization can be traced to adulthood and are manifested in atypical N400 and N170 ERP components (Petrov et al., 2018; Kornilov et al., 2019). It has been shown that adults with a history of institutionalization display reduced neural sensitivity to violations of word expectancy. The results suggest that language is a vulnerable domain in adults with a history of institutionalization, the deficits in which are not explained by general developmental delays and point to the pivotal role of the early linguistic environment in the development of the neural networks involved in language processing. No study to the best of our knowledge has considered very early stages of language processing in children raised in institutions.

The ability to extract native phonological patterns is one of the key components of language development. Studies have demonstrated that infants have an increased general sensitivity, being able to successfully discriminate between sounds of native and non-native languages, gradually becoming attuned to native language and reaching "perceptual narrowing" by the age of 12 months. Perceptual narrowing is an adaptive mechanism that helps to filter out irrelevant linguistic input through perceptual bias (Lewkowicz and Ghazanfar, 2009; Maurer and Werker, 2014). Importantly, the timing of perceptual narrowing can be extended by a number of factors including gestational age (Peña et al., 2012), maternal mental health (Weikum et al., 2012), diet (Innis et al., 2001), and bilingualism (Burns et al., 2007).

Perceptual narrowing has been commonly studied using neuroimaging techniques, including event-related brain potentials (Cheour et al., 2000; Kuhl, 2004), such as the mismatch negativity (MMN) component (Näätänen, 2003). This component is elicited in response to violations of expectation (Winkler, 2007) and has been widely studied as a neural

correlate of phonological discrimination in response to changes in auditory stimulation (Duncan et al., 2009). It plays a pivotal role in speech perception; smaller amplitudes of the MMN component are assumed to reflect poorer speech-sound representations, and as language skills improve, MMN to speech sound contrasts to that language are enhanced (Winkler et al., 1999; Wible et al., 2004). The MMN component can be elicited even in the absence of a participant's attention (Rivera-Gaxiola et al., 2005), and therefore has been widely used in studies with pediatric samples. It has been shown to be sensitive to speech-language and reading difficulties, which are characterized by the altered amplitude of this component compared to typically developing peers (Baldeweg et al., 1999; Cheour et al., 2000; Friederici et al., 2002; Leppänen et al., 2012; Neuhoff et al., 2012; van Zuijen et al., 2013).

Given the impoverished characteristics of the linguistic environment of IC (Windsor et al., 2007; Scott et al., 2011), we hypothesize that children raised in orphanages might demonstrate atypical phonological processing manifested in the discrimination of non-native language patterns after the age of 12 months due to the lack of social interactions in psychosocially depriving environments of institutions.

PARTICIPANTS

A total of 130 children were recruited for the study. However, a number of children ($n = 22$) were excluded according to strict exclusion/inclusion criteria: (1) inability to provide at least 180-30-30 trials to Standard after Standard and Deviant stimuli accordingly ($n = 13$); (2) presence of medically recorded hearing problems ($n = 1$); or (3) diagnosed neurological disorder or neurological symptoms such as epilepsy, brain ischemia, or prenatal brain injury ($n = 7$). One participant was excluded due to previous exposure to the Hindi language; all other participants were Russian native speakers with no previous exposure to the Hindi language.

We inspected the age distribution among the remaining 108 children and identified outliers who were older than 21 months. Due to the unequal distribution of older children in the IC and biological-family-care (BFC) groups, we excluded observations of children who were older than 21 months of age ($n = 26$). The final dataset included ERP data from 82 participants. They were split into two age groups according to the age of hypothesized perceptual narrowing (Rivera-Gaxiola et al., 2005; Maurer and Werker, 2014): the younger age group before 12 months and the older age group after the age of 12 months.

The younger age group included 17 children from the IC group ($M = 10.5$ months, $SD = 1.18$, 11 males) from four baby homes, and 17 children from the BFC group ($M = 10.1$ months, $SD = 1.09$, 12 males). The older group consisted of 23 children from the IC group ($M = 17$ months, $SD = 2.26$, 11 males), and 25 children from the BFC group ($M = 16.9$ m, $SD = 2.25$, 13 males). The groups did not differ significantly by age or sex distribution.

Written consent for participation was obtained from the children's official representatives, baby home officials or

biological parents. The study procedure was approved by the Institutional Review Board (Ethical Committee) of Saint Petersburg State University, Russia.

METHOD

To elicit the MMN ERP component, we used a passive double oddball paradigm (Conboy and Kuhl, 2011). Stimuli were comprised of stop consonant-vowel syllables. We used the /d̪u:/ syllable as a standard stimulus, and /gu:/ and /d̪u:/ as the deviants. Standard /du:/ and deviant /gu:/ were classified as native language patterns; the deviant /d̪u:/ was classified as a foreign phonological pattern from Hindi. The experiment consisted of 1,500 trials, with 1,200 standard (d̪u:/) and 300 deviant (150 /gu:/ and 150 /d̪u:/) trials in total, therefore the ratio of standard to deviant syllables was 8:1:1 (Table 1).

Trials were split into three blocks with 500 stimuli each. Brief 5-min breaks were given between the trial blocks. The stimuli were recorded by a female native Hindi speaker using PRAAT audio software at a sample rate of 44,100 Hz, and presented at 70 dB (SPL) using a set of Yamaha NS-BP300 speakers. Stimuli were administered in a pseudo-randomized order to allow for at least three standard stimuli between deviants; the inter-stimulus interval was 600 ms.

PROCEDURE

The EEG signal was detected using a high-density EEG system via a PC laptop running PyCorder software (BrainProducts Inc.). Specifically, we used the actiCHamp amplifier (BrainProducts, Inc.) to record EEG from the scalp using 64 Ag/AgCl sintered active electrodes mounted in an elastic cap according to the standard montage using SuperVisc electrolyte gel. The signal was recorded using linked mastoids as the reference and digitized at 1,000 Hz.

Data of 31 participants were recorded with online filter settings of 0.10–30 Hz and data of 51 participants were obtained with online filter settings of 0.10–50 Hz. An additional notch filter at 50 Hz was applied to the data online. This inconsistency in data acquisition was attributed to a violation of the research protocol, which was handled at the preprocessing step.

All impedances were kept below 25 kΩ. During the recording, children sat on a caregiver's lap and watched a muted cartoon on a laptop, while auditory stimuli were presented through open field speakers binaurally. Caregivers were instructed not to attend and/or react to stimuli. The EEG data were processed offline using BrainVision Analyzer software v 2.1 (BrainProducts Inc.). The signal was downsampled to 500 Hz. After visual inspection of the raw data for each participant, channels contaminated by noise were reconstructed using spherical spline interpolation.

The signal was re-referenced to the common average reference. IIR filter (low cut-off: 0.10; high cut-off: 30 Hz) was applied to the signal in order to homogenize the filter settings across all participants, followed by a 50 Hz notch filter. We used Independent Component Analysis (ICA) to perform the ocular correction procedure. One of the frontal electrodes (FP1 or FP2 depending on the quality of the recoding) served as a blink marker channel for vertical activity. The difference between FP9 and FP10 electrodes served as a marker for horizontal activity. The Infomax algorithm was trained on a segment of data with a length of 140 s. The procedure was conducted in the semi-automatic mode. After the ICA matrix was computed, the ICA components were visually inspected for each participant with regard to their topographic location and relative impact on the data. The components that were contributing to blinks were set to zero. In total, a maximum number of five ICA components were set to zero for each participant.

After that data was segmented into epochs with 100 ms prestimulus (served as baseline) and 700 ms poststimulus intervals, semi-automatic artifact rejection was carried out. The criteria for artifact rejection were: a voltage step of no more than 50 μV in the segment; and an absolute voltage not exceeding ±110 μV in any of the EEG channels. Baseline correction was performed in relation to the prestimulus time mentioned above and local DC detrending was applied to the extracted segments. The segments were averaged separately for the three experimental conditions: Standard, Native Deviant, Foreign Deviant. Trials in which a Standard stimulus directly followed a Native Deviant/Foreign Deviant were not used in the analysis. Participants were administered different numbers of trials, depending on their distress level and functional state, with minimum of 716 and maximum of 1,500 trials. During the artifact rejection procedure, trials containing exceeding amounts of noise were removed from the analysis (number of removed trials ranged from 77 to 480 segments for each participant, $M = 281.03$, $320 SD = 192.26$). Therefore, on average 637.07 trials for Standard condition were retained ($min = 421$, $max = 877$, $SD = 126.36$); 107.03 Native deviants ($min = 74$, $max = 148$, $SD = 21.27$) and 106.47 Foreign Deviants trials ($min = 72$, $max = 247$, $SD = 21.18$) were left after the artifact rejection.

RESULTS

First, we conducted a *t*-test to ascertain whether the grand average waveforms of Deviant and Standard stimuli significantly differed from zero. All 64 channels were included in the grand average waveforms. There was a significant effect for all experimental conditions, suggesting that a comparison of electric brain activity in response to different experimental conditions

TABLE 1 | Types of auditory stimuli in the event-related potential (ERP) experiment.

Stimulus type	Syllable	Language, where the pattern is present	# of trials	Properties	Duration
Standard	/du:/	Russian and Hindi	1,200	Voiced dental	246 ms
Foreign deviant	/d̪u:/	Hindi	150	Retroflex	242 ms
Native deviant	/gu:/	Russian and Hindi	150	Voiced velar	246 ms

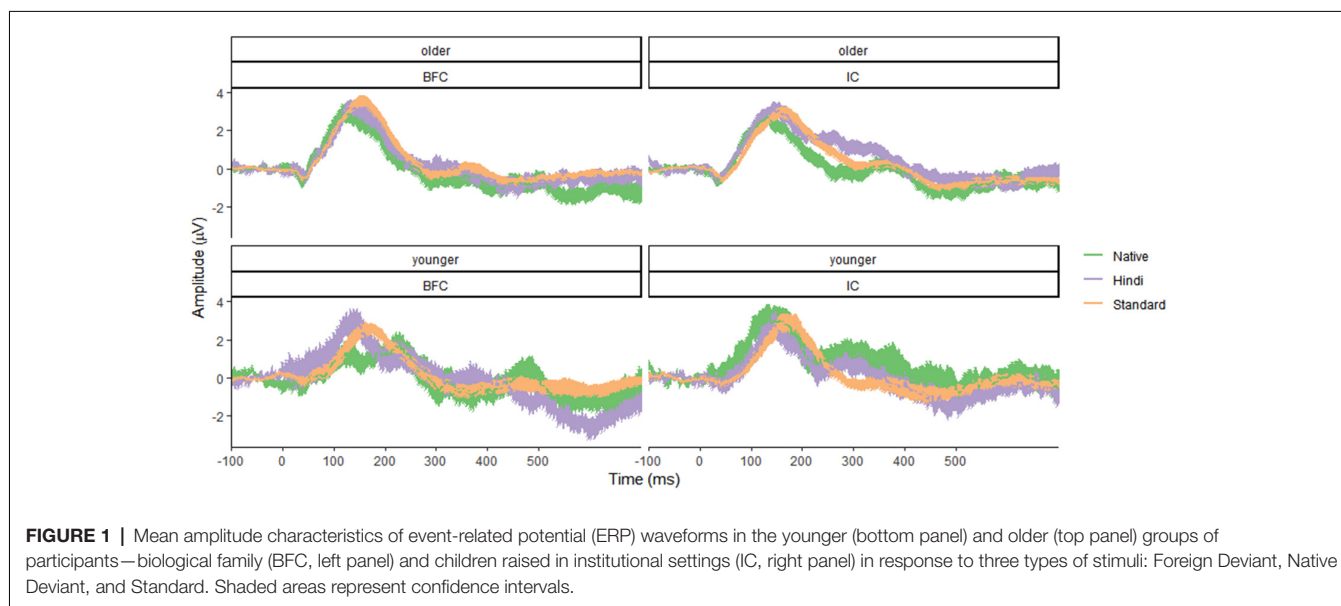
is meaningful. To identify the best time window for the MMN analysis two difference waveforms were computed: Native Deviant—Standard, and Foreign Deviant—Standard. A *t*-test was conducted to compare whether the computed difference waveforms significantly differed from 0, suggesting the presence of MMN effect. For the difference waveform between the Native Deviant and the Standard, statistically significant effect was found in the time window of 125–225 ms after stimulus onset. No significant effect for the difference waveform of the Foreign Deviant and the Standard was observed. Since a significant difference between Native Deviant and Standard conditions was found in the time window of 125–225 ms after the stimulus onset, this latency range was selected as the time window for subsequent analysis.

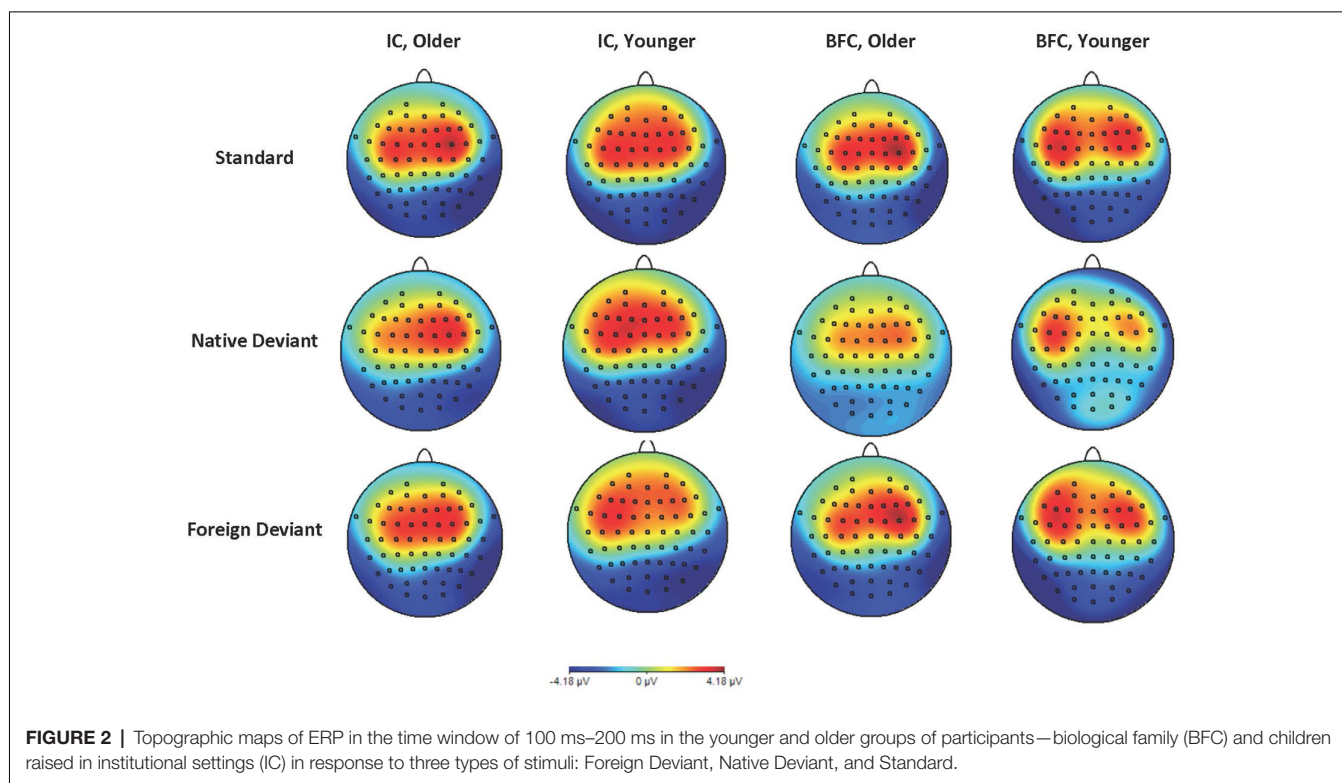
MMN is a component that is observed in the fronto-central electrode sites (Duncan et al., 2009), therefore we first focused our analysis on Left fronto-central (F3, FC5, C3, CP5, F5, C5, CP3, FC3), Midline fronto-central (FC1, Fz, CP1, CP2, Cz, FC2, AF3, AFz, F1, FCz, C1, C2, CPz, F2, AF4), and Right fronto-central (CP6, C4, FC6, F4, C6, FC4, F6, CP4) electrode sites. The younger and older groups of children were analyzed separately to account for potential differences in phonological processing due to perceptual specialization that occurs after the age of 12 months (Kuhl, 2004). In the younger age group, there was no significant effect of electrode cluster in predicting average amplitude differences across experimental conditions ($F_{(2,288)} = 1.75$, $p = 0.17$, Cohen's $f = 0.11$), however in the older age group a significant effect of electrode cluster was found ($F_{(2,414)} = 14.09$, $p < 0.001$, $f = 0.26$). To account for those differences and to keep subsequent statistical analysis consistent across the age groups we moved to individual electrode analysis. The average amplitude in the Fz electrode was selected as an outcome variable in line with previous research (Näätänen et al., 2004; Bishop, 2007).

We utilized a factorial ANOVA to compare the main effects of group (IC/BFC) and stimulus type (Standard, Native Deviant, Foreign Deviant), as well as an interaction effect between group and stimulus type, using the average amplitude of the frontal central electrode (Fz) as an outcome variable. We calculated the mean amplitude for each participant and type of stimulus separately. Statistical analysis was conducted in each age group separately. Tukey correction for multiple comparison was used to correct for the number of experimental conditions in the analysis. Alpha level was 0.05.

Results for the younger age group showed no significant effects of group ($F_{(1,96)} = 1.37$, $p = 0.24$, $f = 0.12$), stimulus type ($F_{(2,96)} = 0.82$, $p = 0.45$, $f = 0.13$), or their interaction ($F_{(2,96)} = 0.73$, $p = 0.49$, $f = 0.12$), suggesting that no MMN effect was registered. The group effect was not significant, indicating the absence of any significant differences between the IC and BFC groups in response to the auditory stimuli in the younger age group. Results for the older age group demonstrated that the type of stimulus effect was significant ($F_{(2,138)} = 3.695$, $p = 0.027$, $f = 0.23$), with greater negativity in response to the Native Deviant stimuli compared to the Standard stimuli [$M = -1.02$, $p = 0.04$, 95% CI (-2.02 , -0.03)]. No significant differences were found between the Standard and Foreign Deviants ($p = 0.977$), as well as between the Native and Foreign Deviants ($p = 0.067$). The group effect was not significant for the older age group as well ($f = 0.05$), indicating the absence of any significant differences in phonological processing between the IC and BFC (Figures 1, 2; Supplementary Tables S1, S2). Also, there was no interaction effect between stimulus type and group factor ($f = 0.04$).

A *post hoc* power analysis revealed that we had 64% power to detect an effect size of $f = 0.2$ in the older age group, given the sample size of 48 children. In the younger age group ($n = 34$) we had 60% power to detect an effect of $f = 0.2$. We believe that the modest sample sizes in each group may have





played a role in our inability to detect the significance of the statistical comparisons conducted, in particular in the younger age group.

DISCUSSION

Our results demonstrate the absence of any MMN effect in the younger age group from our sample, which contradicts the findings presented in the literature. Previously it has been shown that children before the age of 12 months have sensitivity to native, as well as foreign phonological patterns (Maurer and Werker, 2014), therefore we expected to see an MMN effect in the younger age group for the Native Deviant and Foreign Deviant stimuli. The absence of any MMN effect in the younger age group can be explained by the heterogeneity of this component in pediatric samples. It has been shown that the amplitude and polarity of the MMN component changes as a function of age (Friederici et al., 2002; Kushnerenko et al., 2002). Since in this analysis we have used average amplitude of electrical brain activity as an outcome variable, we hypothesize that the MMN effect could be attenuated due to averaging. Also considering the modest sample size, the study could be underpowered for detecting significant results.

In line with our prediction, our findings show that children in the older group are sensitive to native deviants and do not differentiate foreign language contrasts. These findings are in correspondence with the existing literature, which describes perceptual narrowing and reduced sensitivity to non-native language contrasts in typically developing children after the age

of 12 months (Werker and Tees, 1984; Cheour et al., 1998; Rivera-Gaxiola et al., 2005). Specifically, Cheour et al. (1998) reported that infants at 6 months showed a discriminatory response to both native and non-native vowel stimuli, but that by the age of 12 months neural responses to the non-native vowel contrasts were attenuated. A study that also used Hindi non-native deviant consonants showed that children at 7 months of age reveal discrimination of both native and non-native phonetic contrasts, and lose sensitivity to non-native contrasts by the age of 11 months (Rivera-Gaxiola et al., 2005). In addition, we have replicated previous findings that suggest that the MMN effect is observed for native but not foreign language contrasts in typically developing children.

Contrary to our prediction, there was no significant group effect of institutional vs. family environment, indicating that children in the IC group, similar to typically developing peers in biological families, are not sensitive to foreign language contrasts without prolonged exposure to the foreign language. Our initial hypothesis posited that given the impoverished linguistic input in baby homes, children in IC would demonstrate sensitivity to foreign language patterns after the age of 12 months, revealing poorer phonetic representations and discrimination skills. This hypothesis was rejected as the data indicate the presence of significant stimulus type effect for native but not foreign deviants in the older age group compared to the standard stimulus for all children, regardless of their living environment. This study was one of the first attempts to investigate the neural processes underlying the language development of children in institutions using ERP.

Previous studies have demonstrated that children raised in IC demonstrate poor sentence comprehension (Desmarais et al., 2012), low scores in the expressive language domain coupled with hypoactivation of the Broca area (Helder et al., 2014), as well as structural changes and white matter abnormalities in brain areas associated with language, such as the left superior longitudinal fasciculus (Govindan et al., 2010) and arcuate fasciculus (Kumar et al., 2014).

Our study aimed to extend the existing literature by providing data on an intermediate language phenotype in IC children. We aimed to analyze preattentive lower-level language processing characteristics, thus choosing MMN as the component of interest. Our study suggests that the discriminability of auditory information is intact in children raised in institutions, opening up questions regarding the higher-order mechanisms that might explain language deficits in IC children. Thus, based on recent theoretical views of perception narrowing in general and the MMN component in particular as stages in the formation of prediction (and prediction error) in language processing (Bornkessel-Schlesewsky and Schlesewsky, 2019), it will be important to interrogate the IC-BFC group differences in other language-related negative ERP components (e.g., the LAN and N400).

The majority of studies on the MMN component published in Russia have used it as a marker of cognitive decline in various conditions, including stroke (Garin and Poverennova, 2008), dementia (Morozova et al., 2012), schizophrenia (Chepikova et al., 2015; Petrov et al., 2017), and exposure to radiation (Zavoronkova et al., 2010), or as a method of studying attention in typically developing adults (Hodanovich et al., 2009; Gorjainova et al., 2019). Research on Russian children using the MMN component is more scarce. It has been used to study cognitive functions in infants (Vasil'eva et al., 2015) and brain development in children raised in the harsh climatic conditions of the Russian North (Nagornova et al., 2018); also MMN has been proven to be an effective measure for identifying attentional deficits. Moreover, a study using the MMN component established auditory processing deficits in children with motor dysphasia (Savel'eva et al., 2015). No studies published in Russia have considered MMN characteristics in children younger than 3 years of age or children raised in impoverished environments, making this study the first of its kind.

The current study had a number of limitations. First, given the heterogeneity of the MMN component (in terms of spatial distribution and amplitude polarity across developmental milestones; Bishop, 2007), the current sample size might not have been large enough to yield adequate statistical power. Second, the Foreign deviant stimuli were shorter in duration compared to the Standard and Native deviants. These durations should be considered in designing future studies, however, this aspect is unlikely to affect the results, as we observed no significant differences in the responses to Foreign deviants compared to Standard stimuli. Third, it has been reported that MMN amplitude is related to the amount of speech exposure (Marklund et al., 2019); thus, it is important to explore the specifics of language interaction in the IC group (e.g., the

amount of received and produced speech by a child), which, to our knowledge, has never been done. Finally, there are multiple MMN paradigms—e.g., whole word storage MMN, syntactic MMN (Hanna et al., 2017)—we utilized only one, which limits the generalizability of our conclusions. Finally, the auditory stimuli were presented through open field speakers, and caregivers were not wearing sound-canceling headphones. Even though they were instructed not to attend to stimuli, the study does not control for potential caregiver's impact on child attention to the stimuli.

Future studies should continue to interrogate the mechanics of the observed language deficits in individuals who have experienced early institutionalization by extending the MMN paradigm to include other types of stimuli and exploring neurobiological components related to higher-level language processing. In this way, potential biomarkers of language problems in the subpopulation of institutionalized children may be identified.

DATA AVAILABILITY

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

Written informed consent for participation was obtained from the children's official representatives, baby home officials or parents. The study procedure was approved by the Institutional Review Board (Ethical Committee) of Saint Petersburg State University, Russia.

AUTHOR CONTRIBUTIONS

EG led the development and conceptualization of the overall research effort. MV contributed to the study and stimuli design. IO, MZ, AL, and MV collected the data. IO, MZ, AL, and MP preprocessed the EEG data and performed the statistical data analysis. IO, MZ, and EG drafted the first version of the manuscript. IO and MZ prepared the figures.

FUNDING

This research was supported by grant No 14.Z50.31.0027 from the Government of the Russian Federation (PI: EG).

ACKNOWLEDGMENTS

We are grateful to Sergey A. Kornilov for his essential contribution to the study and stimuli design and the data acquisition platform, and to Mei Tan for her editorial support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnhum.2019.00300/full#supplementary-material>

REFERENCES

- Baldeweg, T., Richardson, A., Watkins, S., Foale, C., and Gruzelier, J. (1999). Impaired auditory frequency discrimination in dyslexia detected with mismatch evoked potentials. *Ann. Neurol.* 45, 495–503. doi: 10.1002/1531-8249(199904)45:4<495::aid-ana11>3.0.co;2-m
- Bishop, D. (2007). Using mismatch negativity to study central auditory processing in developmental language and literacy impairments: where are we and where should we be going? *Psychol. Bull.* 133, 651–672. doi: 10.1037/0033-2909.133.4.651
- Bornkessel-Schlesewsky, I., and Schlesewsky, M. (2019). Toward a neurobiologically plausible model of language-related, negative event-related potentials. *Front. Psychol.* 10:298. doi: 10.3389/fpsyg.2019.00298
- Burns, T. C., Yoshida, K. A., Hill, K., and Werker, J. F. (2007). The development of phonetic representation in bilingual and monolingual infants. *Appl. Psycholinguist.* 28, 455–474. doi: 10.1017/s0142716407070257
- Cheour, M., Ceponiene, R., Lehtokoski, A., Luuk, A., Allik, J., Alho, K., et al. (1998). Development of language-specific phoneme representations in the infant brain. *Nat. Neurosci.* 1, 351–353. doi: 10.1038/1561
- Cheour, M., Leppänen, P. H., and Kraus, N. (2000). Mismatch negativity (MMN) as a tool for investigating auditory discrimination and sensory memory in infants and children. *Clin. Neurophysiol.* 111, 4–16. doi: 10.1016/s1388-2457(99)00191-1
- Chepikova, K. A., Mikushkina, L. V., and Gorbunov, I. A. (2015). Osobennosti psihofiziologicheskikh mekhanizmov prognozirovaniya posledovatel'nostej sobytij pri shizofrenii [Characteristics of psychophysiological mechanisms of forecasting event sequences in patients with schizophrenia]. *Aktual. Napravl. Nauch. Issled. Praktike* 3, 232–245.
- Conboy, B. T., and Kuhl, P. K. (2011). Impact of second-language experience in infancy: brain measures of first- and second-language speech perception. *Dev. Sci.* 14, 242–248. doi: 10.1111/j.1467-7687.2010.00973.x
- Desmarais, C., Roeber, B. J., Smith, M., and Pollak, S. (2012). Sentence comprehension in postinstitutionalized school-age children. *J. Speech Lang. Hear. Res.* 55, 45–54. doi: 10.1044/1092-4388(2011)10-0246
- Duncan, C. C., Barry, R. J., Connolly, J. F., Fischer, C., Michie, P. T., Näätänen, R., et al. (2009). Event-related potentials in clinical research: guidelines for eliciting, recording, and quantifying mismatch negativity, P300 and N400. *Clin. Neurophysiol.* 120, 1883–1908. doi: 10.1016/j.clinph.2009.07.045
- Eigsti, I. M., Weitzman, C., Schuh, J., de Marchena, A., and Casey, B. J. (2011). Language and cognitive outcomes in internationally adopted children. *Dev. Psychopathol.* 23, 629–646. doi: 10.1017/S0954579411000204
- Friederici, A. D., Friedrich, M., and Weber, C. (2002). Neural manifestation of cognitive and precognitive mismatch detection in early infancy. *Neuroreport* 13, 1251–1254. doi: 10.1097/00001756-200207190-00006
- Garin, D. P., and Poverennova, I. E. (2008). Izmeneniya akusticheskikh kognitivnykh vyzvannykh potencialov (P300) pri hronicheskoy ishemii mozga [Assessment of auditory cognitive event-related potentials (P300) in patients with chronic stroke]. *Sarat. Nauch. Med. Zhurnal* 4, 86–90.
- Gorjainova, G. Y., Shestakova, A. N., Osadchij, A. E., and Shtyrov, Y. Y. (2019). Otrazhenie avtomaticheskoy obrabotki leksicheskoy informacii v zritel'nykh vyzvannykh potencialah mozga cheloveka [Signs of automatic processing of lexical information in visual brain event-related potentials]. [Elektronnyj Resurs] // *Mezhdunarodnyj Simpozium po Nejroimidzhingu: Fundamental'nye Issledovaniya i Klinicheskaja Praktika*. Available online at: http://psyjournals.ru/neuroimag_2012/issue/53713.shtml. Accessed March 18, 2019.
- Govindan, R. M., Behen, M. E., Helder, E., Makki, M. I., and Chugani, H. T. (2010). Altered water diffusivity in cortical association tracts in children with early deprivation identified with tract-based spatial statistics (TBSS). *Cereb. Cortex* 20, 561–569. doi: 10.1093/cercor/bhp122
- Hanna, J., Cappelle, B., and Pulvermüller, F. (2017). Spread the word: MMN brain response reveals whole-form access of discontinuous particle verbs. *Brain Lang.* 175, 86–98. doi: 10.1016/j.bandl.2017.10.002
- Helder, E. J., Behen, M. E., Wilson, B., Muzik, O., and Chugani, H. T. (2014). Language difficulties in children adopted internationally: neuropsychological and functional neural correlates. *Child Neuropsychol.* 20, 470–492. doi: 10.1080/09297049.2013.819846
- Hodanovich, M. Y., Esipenko, E., Svetlik, M. V., and Krutenkova, E. P. (2009). Zritel'nyj analog negativnosti rassoglasovanija pri otklonenii stimulov po dlitel'nosti [Visual analogue of mismatch negativity in stimuli with deviant duration]. *Zhurnal Vysshej Nervnoj Dejatel'nosti im. I.P. Pavlova* 59, 296–306.
- Innis, S. M., Gilley, J., and Werker, J. (2001). Are human milk long-chain polyunsaturated fatty acids related to visual and neural development in breast-fed term infants? *J. Pediatr.* 139, 532–538. doi: 10.1067/mpd.2001.118429
- Kornilov, S. A., Zhukova, M. A., Ovchinnikova, I. V., Golovanova, I. V., Naumova, O. Y., Logvinenko, T. I., et al. (2019). Language outcomes in adults with a history of institutionalization: behavioral and neurophysiological characterization. *Sci. Rep.* 9:4252. doi: 10.1038/s41598-019-40007-9
- Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nat. Rev. Neurosci.* 5, 831–843. doi: 10.1038/nrn1533
- Kuhl, P. K., Tsao, F.-M., and Liu, H.-M. (2003). Foreign-language experience in infancy: effects of short-term exposure and social interaction on phonetic learning. *Proc. Natl. Acad. Sci. U S A* 100, 9096–9101. doi: 10.1073/pnas.1532872100
- Kumar, A., Behen, M. E., Singsoonsud, P., Veenstra, A. L., Wolfe-Christensen, C., Helder, E., et al. (2014). microstructural abnormalities in language and limbic pathways in orphanage-reared children: a diffusion tensor imaging study. *J. Child Neurol.* 29, 318–325. doi: 10.1177/0883073812474098
- Kushnerenko, E., Ceponiene, R., Balan, P., Fellman, V., and Näätänen, R. (2002). Maturation of the auditory change detection response in infants: a longitudinal ERP study. *Neuroreport* 13, 1843–1848. doi: 10.1097/00001756-200210280-00002
- Leppänen, P., Hämäläinen, J., Guttorm, T., Eklund, K., Salminen, H., Tanskanen, A., et al. (2012). Infant brain responses associated with reading-related skills before school and at school age. *Neurophysiol. Clin.* 42, 35–41. doi: 10.1016/j.neucli.2011.08.005
- Lewkowicz, D. J., and Ghazanfar, A. A. (2009). The emergence of multisensory systems through perceptual narrowing. *Trends Cogn. Sci.* 13, 470–478. doi: 10.1016/j.tics.2009.08.004
- Loman, M., Wiik, K., Frenn, K., Pollak, S., and Gunnar, M. (2009). Postinstitutionalized children's development: growth, cognitive, and language outcomes. *J. Dev. Behav. Pediatr.* 30, 426–434. doi: 10.1097/dbp.0b013e3181b1fd08
- Marklund, E., Schwarz, I.-C., and Lacerda, F. (2019). Amount of speech exposure predicts vowel perception in four- to eight-month-olds. *Dev. Cogn. Neurosci.* 36:100622. doi: 10.1016/j.dcn.2019.100622
- Maurer, D., and Werker, J. F. (2014). Perceptual narrowing during infancy: a comparison of language and faces. *Dev. Psychobiol.* 56, 154–178. doi: 10.1002/dev.21177
- Morozova, A., Evtushenko, S., and Morozova, T. (2012). Mul'tifokal'nye vyzvannye potencialy, svyazannye s sobytijem, v rannej diagnostike kognitivnoj dezintegracii: kliniko-nejrofiziologicheskaja traktovka [Multifocal event related potentials in early diagnostics of cognitive decline: clinical and neurophysiological context]. *Mezhdunarodnyj Nevrologicheskij Zhurnal* 3, 26–41.
- Muhamedrahimov, R. J., Agarkova, V. V., Vershnina, E. A., Palmov, O. I., Nikiforova, N. V., Mccall, R. B., et al. (2014). Behavior problems in children transferred from a socioemotionally depriving institution to St. Petersburg (Russian Federation) families. *Infant Mental Health J.* 35, 111–122. doi: 10.1002/imhj.21435
- Muhamedrahimov, R. J., Nikiforova, N. V., Palmov, O. I., Groark, C. J., McCall, R. B., and Larry, F. (2005). Characteristics of children, caregivers, and orphanages for young children in St. Petersburg, Russian Federation. *J. Appl. Dev. Psychol.* 5, 477–506. doi: 10.1016/j.appdev.2005.06.002
- Näätänen, R. (2003). Mismatch negativity: clinical research and possible applications. *Int. J. Psychophysiol.* 48, 179–188. doi: 10.1016/s0167-8760(03)00053-9
- Näätänen, R., Pakarinen, S., Rinne, T., and Takegata, R. (2004). The mismatch negativity (MMN): towards the optimal paradigm. *Clin. Neurophysiol.* 115, 140–144. doi: 10.1016/j.clinph.2003.04.001
- Nagornova, Z. V., Shemyakina, N. V., Soroko, S. I., and Belisheva, N. K. (2018). Analiz vozrastnoj dinamiki i polovykh osobennostej spontannoj bioelektricheskoy aktivnosti i komponentov sluhovykh vyzvannykh potencialov u mladshih shkol'nikov, prozhivajushhih v Arkticheskoy zone RF [Analysis

- of Age-Related Dynamics and Gender-Specific Characteristics of Spontaneous Bioelectrical Activity and Components of Auditory Evoked Potentials in Junior School Students Living in the Arctic Region of the Russian Federation]. *Fiziologija Cheloveka* 44, 84–95.
- van Zuijlen, T. L., Plakas, A., Maassen, B. A., Maurits, N. M., and van der Leij, A. (2013). Infant ERPs separate children at risk of dyslexia who become good readers from those who become poor readers. *Dev. Sci.* 16, 554–563. doi: 10.1111/desc.12049
- Neuhoff, N., Bruder, J., Bartling, J., Warnke, A., Remschmidt, H., Müller-Myhsok, B., et al. (2012). Evidence for the late MMN as a neurophysiological endophenotype for dyslexia. *PLoS One* 7:e34909. doi: 10.1371/journal.pone.0034909
- Peña, M., Werker, J. F., and Dehaene-Lambertz, G. (2012). Earlier speech exposure does not accelerate speech acquisition. *J. Neurosci.* 32, 11159–11163. doi: 10.1523/JNEUROSCI.6516-11.2012
- Petrov, M. V., Golovanova, I., Ovchinnikova, I., Logvinenko, T., Zhukova, M., Kornilov, S., et al. (2018). Left-lateralized N170 effects in reading of words with high and low frequency in adults with a history of institutionalization. *Int. J. Psychophysiol.* 131, S51–S52. doi: 10.1016/j.ijpsycho.2018.07.155
- Petrov, M. V., Kolchev, A. I., Ershov, B. B., Gvozdeckij, A. N., Golovanova, I. V., and Daeva, N. A. (2017). Vzaimosvjaz' redukcii negativnosti rassoglasovanija i kognitivnogo deficita pri paranoidnoj shizofrenii [Association between mismatch negativity and cognitive deficit in patients with paranoid schizophrenia]. *Vestnik Sankt-Peterburgskogo Universiteta. Serija 16. Psihologija. Pedagogika* 7, 91–103.
- Rivera-Gaxiola, M., Klarman, L., Garcia-Sierra, A., and Kuhl, P. K. (2005). Neural patterns to speech and vocabulary growth in American infants. *Neuroreport* 16, 495–498. doi: 10.1097/00001756-200504040-00015
- Savel'eva, N. A., Anisimov, G. V., and Kalashnikova, T. P. (2015). Pokazateli kognitivnyh vyzvannyh potencialov u detej s rechevym dizontogenezom [Characteristics of cognitive event-related potentials in children with developmental language disorder]. *Fundamental'nye Issledovanija* 2, 346–349.
- Scott, K. A., Roberts, J. A., and Glennen, S. (2011). How well do children who are internationally adopted acquire language? A meta-analysis. *J. Speech Lang. Hear. Res.* 54, 1153–1169. doi: 10.1044/1092-4388(2010/10-0075)
- Vasil'eva, M. J., Bannyh, O. N., Koschavtsev, A. G., Vershinina, E. A., and Alexandrov, A. A. (2015). Specific characteristics of mismatch negativity in infants. *Ross. Fiziol. Zh. Im. I. M. Sechenova* 101, 85–97.
- Vorria, P., Ntouma, M., and Rutter, M. (2014). The behaviour of adopted adolescents who spent their infancy in residential group care: the Greek Metera study. *Adopt. Foster.* 38, 271–283. doi: 10.1177/0308575914543237
- Weikum, W. M., Oberlander, T. F., Hensch, T. K., and Werker, J. F. (2012). Prenatal exposure to antidepressants and depressed maternal mood alter trajectory of infant speech perception. *Proc. Natl. Acad. Sci. U S A* 109, 17221–17227. doi: 10.1073/pnas.1121263109
- Werker, J. F., and Tees, R. C. (1984). Cross-language speech perception: evidence for perceptual reorganization during the first year of life. *Infant Behav. Dev.* 7, 49–63. doi: 10.1016/s0163-6383(84)80022-3
- Wible, B., Nicol, T., and Kraus, N. (2004). Atypical brainstem representation of onset and formant structure of speech sounds in children with language-based learning problems. *Biol. Psychol.* 67, 299–317. doi: 10.1016/j.biopsycho.2004.02.002
- Windsor, J., Glaze, L. E., Koga, S. F., and Proje, B. E. I. (2007). Language acquisition with limited input: romanian institution and foster care. *J. Speech Lang. Hear. Res.* 50, 1365–1381. doi: 10.1044/1092-4388(2007/095)
- Winkler, I. (2007). Interpreting the mismatch negativity. *J. Psychophysiol.* 21, 147–163. doi: 10.1027/0269-8803.21.34.147
- Winkler, I., Kujala, T., Tiitinen, H., Sivonen, P., Alku, P., Lehtokoski, A., et al. (1999). Brain responses reveal the learning of foreign language phonemes. *Psychophysiology* 36, 638–642. doi: 10.1017/s0048577299981908
- Zhavoronkova, L., Belostockij, A., Kulikov, M., Oknina, L., Holodova, N., and Kupcova, S. (2010). Osobennosti sluhovyh kognitivnyh vyzvannyh potencialov u uchastnikov likvidacii posledstvij Chernobyľ'skoj avarii. Soobshhenie I. Analiz rannego N1 komponenta [Characteristics of auditory cognitive evoked potentials in liquidators of Chernobyl Nuclear Power Plant accident. Report I. Analysis of the early N1 component]. *Fiziologija Cheloveka* 36, 32–43.
- Zhukova, M., Kornilov, S., Vasilyeva, M., Letunovskaia, A., Muhamedrahimov, R., and Grigorenko, E. (2015). Delayed language development in young children raised in institutional care is manifested in the atypical N400 component. In *Abstracts of the 7th Annual Meeting of the Society for the Neurobiology of Language* (Chicago, IL: Routledge, Taylor & Francis Group), 75–76. Available online at: http://www.neurolang.org/programs/SNL2015_Abtracts.pdf

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Ovchinnikova, Zhukova, Luchina, Petrov, Vasilyeva and Grigorenko. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Novel Word Learning: Event-Related Brain Potentials Reflect Pure Lexical and Task-Related Effects

Beatriz Bermúdez-Margaretto^{1*}, David Beltrán^{2,3}, Fernando Cuetos⁴
and Alberto Domínguez^{2,3}

¹Centre for Cognition and Decision Making, Institute for Cognitive Neuroscience, National Research University Higher School of Economics, Moscow, Russia, ²Instituto Universitario de Neurociencia (IUNE), Tenerife, Spain, ³Facultad de Psicología, Universidad de La Laguna, Tenerife, Spain, ⁴Facultad de Psicología, Universidad de Oviedo, Oviedo, Spain

OPEN ACCESS

Edited by:

Olga V. Shcherbakova,
Saint Petersburg State University,
Russia

Reviewed by:

Laura Barca,
Italian National Research Council
(CNR), Italy
Manuel De Vega,
University of La Laguna, Spain

*Correspondence:

Beatriz Bermúdez-Margaretto
bermudezmargaretto@gmail.com

Specialty section:

This article was submitted to Speech
and Language, a section of the
journal *Frontiers in Human
Neuroscience*

Received: 28 February 2019

Accepted: 20 September 2019

Published: 15 October 2019

Citation:

Bermúdez-Margaretto B, Beltrán D,
Cuetos F and Domínguez A
(2019) Novel Word Learning:
Event-Related Brain Potentials
Reflect Pure Lexical and
Task-Related Effects.
Front. Hum. Neurosci. 13:347.
doi: 10.3389/fnhum.2019.00347

Previous research has pointed out that the combination of orthographic and semantic-associative training is a more advantageous strategy for the lexicalization of novel written word-forms than their single orthographic training. However, paradigms used previously involve explicit stimuli categorization (lexical decision), which likely influence word learning. In the present study, we used a more automatic task (silent reading) to determine the advantage of the associative training, by comparing the brain electrical signals elicited in combined (orthographic and semantic) and single (only orthographic) training conditions. In addition, the learning effect (in terms of similar neurophysiological activity between novel and known words) was also tested under a categorization paradigm, enabling determination of the possible influence of the training task in the lexicalization process. Results indicated that novel words repeatedly associated with meaningful cues showed a higher attenuation of N400 responses than those trained in the single orthographic condition, confirming the higher facilitation in the lexico-semantic processing of these stimuli, as a consequence of semantic associations. Moreover, only when the combined training was carried out in the reading task did novel words show similar N400 responses to those elicited by known words, suggesting the achievement of a similar lexical processing to known words. Crucially, when the training is carried out under a demanding task context (lexical decision), known words exhibited positive enhancement within the N400 time window, contributing to maintaining N400 differences with novel trained words and confounding the outcome of the learning. Such deflection—compatible with the modulation of the categorization-related P300 component—suggests that novel word learning could be influenced by the activation of categorization-related processes. Thus, the use of low-demand tasks arises as a more appropriate approach to study novel word learning, enabling the build-up process of mental representations, which probably depends on pure lexical and semantic factors rather than being guided by categorization demands.

Keywords: novel word learning, lexical decision task, reading task, event-related brain potentials, N400

INTRODUCTION

The development of reading fluency, namely the ability to visually recognize words with an adequate level of accuracy and speed, is essential for correct performance in most of our daily-life activities and critical for academic and professional success. It is accepted that repeated visual experience with novel word-forms allows the reader to evolve from slow, effortful and inaccurate reading, characterized by serial letter-by-letter decoding to automated and skilled reading, in which words are recognized through direct, parallel processing (Share, 1995, 2008). Thus, after a novel word-form has been decoded several times through purely visual experience, a mental representation is built-up in the reader's lexicon, enabling its reading using this whole-word visual strategy (Meyer and Felton, 1999; Coltheart et al., 2001). Therefore, this so-called lexicalization process is crucial for the acquisition of the direct visual recognition of words and, ultimately, for developing fluent reading. However, the specific training which enables the integration of novel word-forms into the reader's lexicon is still under debate.

Some behavioral studies have claimed that the formation of lexical representations is possible after single orthographic training with novel written word-forms, involving just a handful of repeated visual exposures under meaningless conditions, namely in the absence of any association to a semantic reference. Thus, this training is characterized as meaningless and non-associative, in which novel written-word forms are briefly exposed to participants through a short number of visual presentations (ranging from 4 and 10, depending on the study). In particular, these studies obtained the reduction of the length effect between short and long novel word-forms (Ellis et al., 2009; Maloney et al., 2009; Kwok and Ellis, 2015; Kwok et al., 2017; Suárez-Coalla et al., 2016) or an interference effect in the categorization of known words (Bowers et al., 2005; Qiao and Forster, 2013). Both results are taken as indexes of the representation of the novel items in the reader's lexicon. However, contrary arguments can also be found in the literature. For instance, it is argued that such an interference effect is not indicative of the complete lexicalization of these stimuli but of the storage of episodic memory traces for them, which interfere during the categorization of known words (Leach and Samuel, 2007). Accordingly, other studies have shown that only when both orthography and the meaning of novel words are trained, is it possible to observe lexical competition effects between these stimuli and known words, in terms of a reduction of the prime lexicality effect (Qiao et al., 2009; Qiao and Forster, 2013). Therefore, some authors conclude that orthographic training is not enough to ensure the lexicalization of novel word-forms, with effects denoting the acquisition of interfering-episodic memory traces rather than competing-lexical representations after this training.

Nevertheless, given the rapid and dynamic changes that occur in the linguistic system during novel word learning, other measures than those which are behavioral are required to evaluate this process correctly. Thus, magneto- and electroencephalography methodologies, able to track online-processing changes in brain activity, are probably much more

sensitive to assess novel word learning and, particularly, the nature of the neurophysiological mechanisms underlying a specific training with the orthographic or both the orthographic and semantic features of novel words. Regarding the effect of single orthographic training, rather few MEG/EEG studies are focused on neural dynamics during the acquisition of novel surface word-forms, with substantial methodological differences and inconsistent findings among them (Bermúdez-Margaretto et al., 2015, 2018; Partanen et al., 2018). For instance, in a recent MEG study, Partanen et al. (2018) found that massive (~100 repetitions) and unattended, parafoveal exposure to novel written word-forms caused an increase in the early brain activity, at around 100 ms post-stimulus onset. This enhancement, found after only 15 min of exposure with novel words outside the focus of the reader's attention, was considered indicative of the rapid and automatic formation of lexical traces for these stimuli. However, different results have been found under paradigms better resembling the attentive context in which novel written word-forms are usually encountered. Thus, recent EEG research has shown that single orthographic training with novel word-forms enables the formation of memory traces for these stimuli whose nature is probably episodic rather than lexical (Bermúdez-Margaretto et al., 2015), in agreement with some behavioral studies discussed above (Qiao et al., 2009; Qiao and Forster, 2013). Specifically, short (up to six repetitions) visual exposure to novel word-forms in a lexical decision task caused an increase in amplitude in the late positive component (LPC), an ERP component traditionally related to episodic memory processes and recollection of previously presented information from long-term memory (for a review see Rugg and Curran, 2007). Hence, this LPC effect was considered to index the codification and strengthening of episodic memory traces that follow the repeated exposures of these stimuli. Given no modulation in lexical or lexico-semantic related ERP components was found as a consequence of this orthographic training, it was hypothesized that probably both novel word orthography and meaning should be trained in order to better instantiate them as lexical items.

This hypothesis was tested in a second study (Bermúdez-Margaretto et al., 2018), where we conducted a similar lexical decision task in which short orthographic training with novel word-forms (again, six repetitions) was compared to the effect of training both the orthography and the meaning of the stimuli, simultaneously. Thus, novel written word-forms were repeatedly presented in a single orthographic training (namely, a meaningless training condition) or in a combined orthographic/semantic training condition, where novel word-forms were trained through semantic-associative picture-word exposures (namely, a meaningful training condition). Replicating our previous findings, novel word-forms trained in the meaningless, non-associative condition showed an LPC enhancement across repetitions, reflecting the activation of episodic memory process through single orthographic training. Interestingly, a higher facilitation in the lexico-semantic processing of novel words was found when these stimuli were presented in the meaningful, semantic-associative training, reflected in a higher decrease in the N400 amplitudes for these

stimuli in comparison to those trained in the meaningless condition. The modulation of this ERP component, typically related to semantic processing (Kutas and Federmeier, 2011), was taken as an index of the association between novel word-forms and picture-concepts throughout the meaningful training, in line with previous studies training novel words under meaningful conditions (Perfetti et al., 2005; Mestres-Missé et al., 2007; Borovsky et al., 2010; Frishkoff et al., 2010; Batterink and Neville, 2011; Angwin et al., 2014; Bakker et al., 2015). Notably, this advantage of the combined orthographic/semantic training over the single orthographic training had not been observed before, given that no direct comparison between both trainings had been provided before. Therefore, this study confirmed the effect of semantic training going beyond the enhancement of episodic memory processes, enabling the lexico-semantic facilitation of novel word-forms and probably contributing to their lexicalization to a higher extent.

In the above studies, the task used to guarantee stimuli processing during training was the lexical decision task, in which the primary aim is to categorize the upcoming stimuli—both known and novel words—as lexical/non-lexical items. This task forces the discrimination between known and novel words and could thereby facilitate the learning of the novel word-forms. Moreover, the particular semantic-associative training carried out in this task could further influence the learning of these stimuli, given that the preceding picture enabled their prediction and response anticipation. Thus, since the efficient picture-stimulus association ensured the faster and accurate categorization of the stimuli, participants probably followed an associative strategy in order to successfully fulfill the task requirement, leading to a higher facilitation in the processing of these stimuli and consequently lower N400 amplitudes. Therefore, the particular task context in which the semantic-associative training was carried out probably facilitated the development of a strategic-based learning, which, on the other hand, might be only indirectly related to the formation of the novel word as a lexical item. In this regard, it is possible that the N400 effect found in that study was not only reflecting facilitation in the lexico-semantic processing of stimuli but also its categorization during the task. Indeed, perceptual discrimination processes carried out in order to accomplish task requirements (as in this particular task, stimuli categorization) can also be reflected in this time window, as is the case of the P300 component (Polich, 1985, 2004; Picton, 1992).

Semantic processes are, however, considered to be rather automatic, with the access to stimulus meaning occurring in the absence of specific strategy or intention from the reader, although they might be modulated by higher top-down factors, such as temporal attention or task demands (Kiefer, 2008). For instance, automaticity in meaning access is reflected in the masked semantic priming effect, where the target processing is facilitated by a semantically related prime even when it is perceived unconsciously—and hence automatically (Carr and Dagenbach, 1990; Neely, 2012). Brain electrical signals also reflect such automaticity in semantic processing, with reduced N400 amplitudes elicited by targets preceded by semantically related masked primes (Deacon et al., 2000; Kiefer, 2002).

Thus, facilitation in the lexico-semantic processing of novel word-forms could occur even if meaningful associations are carried out in a task involving a more automatic processing of the stimuli, such as a simple reading task.

Reading, besides preventing possible facilitation in word learning caused by categorization, is significantly less demanding than lexical decision since it involves a much more automatic processing of stimuli. Although some attention-demanding processes occur during reading (such as inference making or comprehension monitoring when reading texts), many others are automatic (such as letter identification or lexico-semantic access), particularly if reading of isolated words is considered (Perfetti, 1985; Walczyk, 2000). Indeed, lexico-semantic processes are accessed during this automatic-driven processing task even in the absence of a particular response; this has been evidenced in several studies, with the modulation of N400 when reading words semantically incongruent with the preceding sentence context (Kutas and Hillyard, 1980; Kutas and Van Petten, 1988). Therefore, the semantic-associative training of novel word-forms could facilitate the lexico-semantic processing of these stimuli during a reading task, confirming the advantage of this training for word lexicalization in the absence of confounding categorization effects. Moreover, this task would result in a more appropriate context to study the acquisition of mental traces for novel word-forms, since no other processes beyond those specifically related to word lexicalization—grapheme-to-phoneme decoding—are involved. In this sense, the presence of an N400 effect even with the suppression of categorization demands could indicate the formation of lexico-semantic traces, non-dependent on these processes but probably reflecting pure associative learning as a consequence of the training.

Therefore, the main goal of the present study was to determine whether the advantage of the combined training, over the single orthographic training, could be replicated under a training task free of categorization-confounding responses (silent reading), indicating the effectiveness of the semantic-associative training in novel word learning, or whether such an advantage was a consequence of the specific categorization context of the task (lexical decision). With this purpose, the present study carried out the same training paradigm as implemented before (Bermúdez-Margaretto et al., 2015)—thus, comparing single orthographic vs. orthographic/semantic trainings—but in this case, a silent reading task was used as a training context, instead of a lexical decision task. Importantly, this task shares the same materials, procedure, features of the sampled participants, EEG equipment and preprocessing pipeline as in the previous lexical decision task (for details see “Materials and Methods” section), making both tasks methodologically comparable. In particular, two main questions were separately addressed in this study.

First, we aimed to determine whether the combination of both orthographic and semantic training with novel word-forms facilitates the lexical processing of these stimuli to a higher extent than the single orthographic training, by using a task context in which the learning of the stimuli is not influenced by categorization demands. To address this question, the effect of both training conditions was tested along the silent reading

task, in a similar way as carried out in our previous lexical decision task. We hypothesized that, as found in our previous study using the lexical decision task (Bermúdez-Margaretto et al., 2015), novel word-forms trained in the meaningful, semantic-associative condition in the present reading task will show greater facilitation in their lexico-semantic processing than non-associated stimuli, reflected in higher attenuation of N400 amplitudes. This training effect would indicate that, even in a task in which stimuli are processed automatically, the combination of orthographic and semantic training results in a more advantageous approach for their learning than in the case of the single orthographic training, with the progressive acquisition of meaningful content through associations to picture-concepts.

Additionally, we considered to explore the impact of the meaningful, semantic-associative training on the lexicality effect, namely in the differences between trained novel word-forms and already known words. This lexicality effect was not tested in our previous lexical decision task since that study was mainly focused on disentangling the effect of training novel words in single orthographic and combined conditions. Therefore, testing the N400 lexicality effect in both task contexts would provide further evidence about the acquisition of memory traces for semantically trained stimuli. Indeed, this effect is thought to reflect differences between already lexicalized stimuli and those without mental representations (Forster and Chambers, 1973; Glushko, 1979). Accordingly, previous studies have concluded that the reduction or absence of the N400 lexicality effect after semantic training evidences the achievement of the lexico-semantic status for trained stimuli (Mestres-Missé et al., 2007; Batterink and Neville, 2011; Bakker et al., 2015). Then, to address whether the semantic-associative training would lead to similar lexico-semantic processing between novel and known words and if this would occur to a different extent across tasks, we evaluated the N400 lexicality effect at the end of the meaningful, semantic-associative training in both tasks, the present silent reading and the previous lexical decision. Lexical differences between known and novel word-forms—and hence, a higher N400 lexicality effect—were expected in the lexical decision rather than in the reading task despite the learning, given the forced discrimination between known and novel words. However, a better match between the processing of novel and known words was expected in reading, confirming the formation of lexical, non-categorization-guided memory traces for stimuli trained in this particular task.

MATERIALS AND METHODS

Participants

A group of 25 undergraduate psychology students took part in the present silent reading task for course credits (23 females; mean age of 21.48; SD: 2.04). All of them were native Spanish speakers, had normal or correct-to-normal vision and were right-handed according to the Oldfield's Handedness Inventory (Oldfield, 1971). No psychiatric or neurological disorder was disclosed by any participant. This research was approved by the Ethics Committee of the Psychology Department of the University of Oviedo. Before starting the experimental tasks,

participants received pertinent information about the purpose of the study, the tasks, and their duration. Written informed consent was then received from participants.

Materials

The present silent reading task used the same materials and design as implemented in the previous lexical decision task (see Bermúdez-Margaretto et al., 2015). Hence, the task was divided into six blocks and the same set of 448 stimuli was used. Sixty-four of these stimuli were novel written word-forms (4–7 letter pseudowords, namely meaningless stimuli observing the orthographic and phonotactic Spanish rules, i.e., *pasne*), repeatedly presented from the first to the sixth block of the task. The remaining 384 stimuli were known words (4–7 letter Spanish nouns, i.e., *barba*), presented in sets of 64 stimuli in each task block. Therefore, these stimuli were not repeated but a new set of known words was presented in each task block. The aim of this procedure was to evaluate the lexicality effect in a more natural way, comparing the processing of a stimulus that is new and repeatedly encountered by the reader—and hence becoming familiar—with the processing of a stimulus that is already known and non-repeated. In sum, both tasks were composed of six blocks, each of them containing 128 stimuli, half of them known and the other half novel word-forms.

Additionally, half of the stimuli (both known and novel word-forms) in each task were repeatedly associated with a known concept by means of the previous presentation of a picture of a known object (semantic-associative condition with combined orthographic/semantic training). The other half of the stimuli were preceded by the presentation of a hash mark (#) not related to a known meaning (non-associative condition with single orthographic training). More specifically, known words (nouns) were associated with the corresponding picture of a known object in association with their meaning, maintaining correspondence between concepts represented by pictures and words. Thus, different pictures were presented in association with known words across blocks, whose selection was based on the word's meaning. Regarding novel words, these stimuli were always associated to the same cue (a picture of a known object or hash mark) across repetitions. For this purpose, another set of pictures of known objects was selected (note that, target words for pictures associated to known and to novel words were counterbalanced in their familiarity and imageability). Pictures of known objects were obtained from the Snodgrass and Vanderwart set of pictures (Snodgrass and Vanderwart, 1980) and both pictures and hash marks had similar appearance and dimensions (10 × 15 cm). **Table 1** shows the matching of the experimental stimuli in the main lexical (familiarity, imageability) and sub-lexical (frequency of bigrams and first syllable, number orthographic neighbors) psycholinguistic variables by means of the BuscaPalabras database (Davis and Perea, 2005).

Procedure

First, an electrode cap was mounted on the scalp of participants, in order to record their EEG activity during the silent reading task. Verbal instructions were given to participants before starting the reading task, namely to pay attention and silently

TABLE 1 | Matching means of each psycholinguistic variable through known and novel words compared for the present study (the same materials were used in both tasks).

		Known words Block 1 (associative condition)	Known words Block 6 (associative condition)	Novel words (associative condition)	Non-novel words (associative condition)	Test	p-value
Sub-lexical variables	Bigram frequency (token type)	590.46 (295.84)	598.30 (309.62)	516.28 (262.79)	515.32 (225.54)	$F_{(3,125)} = 0.87$	0.45
	Number of orthographic neighbors	3.37 (4.36)	2.46 (3.07)	2.68 (3.71)	1.31 (2.05)	$F_{(3,125)} = 2.02$	0.11
	First syllable frequency	291.36 (248.36)	255.45 (208.82)	271.45 (274.02)	306.67 (224.30)	$F_{(3,125)} = 0.27$	0.84
Lexical variables	Imageability	5.46 (1.75)	6.15 (0.36)	5.52 (2.16)	-	$F_{(2,95)} = 1.79$	0.17
	Familiarity	5.55 (1.83)	6.09 (0.68)	5.37 (2.14)	-	$F_{(2,95)} = 1.61$	0.20

Standard deviations are shown in brackets. Statistical contrasts confirmed no significant differences across compared conditions (all post hoc contrasts resulted in $p > 0.05$).

read each stimulus presented on the screen. This procedure was similar to that carried out in the previous lexical decision task, in which an explicit categorization of the stimuli was required (for details see Bermúdez-Margaretto et al., 2015). The researcher emphasized that participants should avoid blinks and muscular movements during the task and encouraged them to take breaks after each task block in order to prevent artifacts and fatigue. Before starting the experiment, instructions for the task appeared on the computer screen followed by eight training trials.

Stimuli were displayed in black Verdana 18 point letters (known and novel words) or in black line drawings (pictures and hash marks) over a white background in the center of the screen by means of the E-Prime 2.0 software (Schneider et al., 2002). All trials were presented in randomized order within each task block. The sequence of stimuli presentation in the current reading task was identical to that of the lexical decision task employed by Bermúdez-Margaretto et al. (2015). In particular, the sequence started with a fixation cross displayed in the center of the screen for 1,000 ms. Then, a picture (for semantic-associative trials) or a hash mark (for non-associative trials) was presented for 150 ms, followed by a 200 ms blank screen. Afterward, the target (a known or a novel word) was presented on the screen for 700 ms (or until participant's response, for the lexical decision task). Finally, another blank screen was presented for 500 ms; see **Figure 1** for the sequence of stimuli presentation in both tasks.

Recording and Pre-processing of the EEG Data

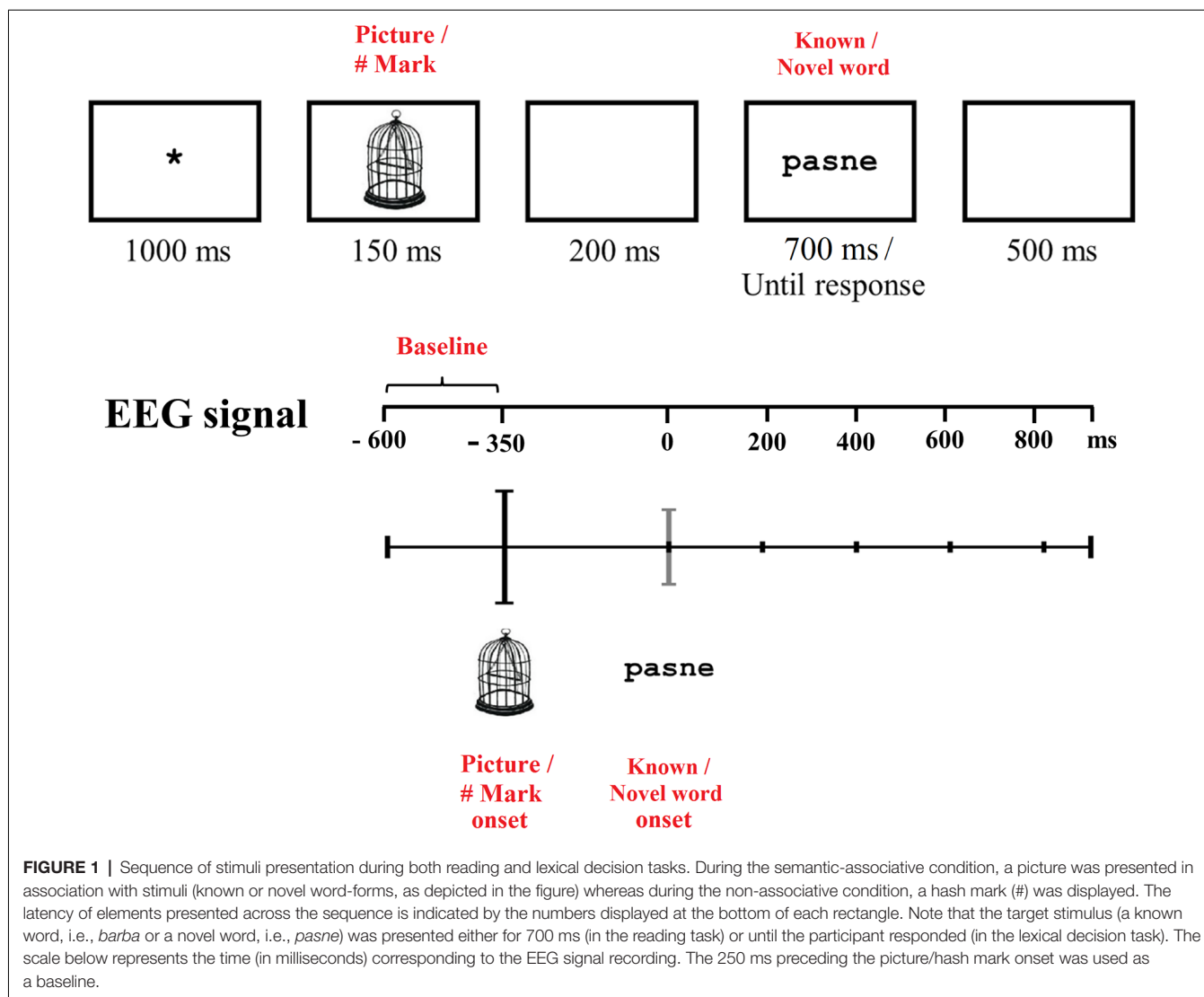
Brain electrical signals were recorded during the present reading task by means of an EEG equipment with 64 Ag/AgCl actiCAP electrodes (Brain Products GmbH, Gilching), similarly to that used in the previous lexical decision task, mounted in an elastic cap according to the 10/20 system (Jasper, 1958). The inter-electrode impedance of active electrodes was kept under 25 k Ω . Ocular activity was recorded by two electrodes placed on the infraorbital and supraorbital canthus of the left eye. The activity in both mastoid bones was also recorded to calculate an offline reference. During the online recordings, the EEG signal was referenced to the activity of the vertex electrode (Cz). The EEG and EOG signals were digitalized and amplified by an actiChamp amplifier system (Brain Products GmbH, Gilching) at a 1,000 Hz

sampling rate. A notch filter at 50 Hz was applied and 0.1 and 100 Hz high and low pass filters were set.

Pre-processing of EEG signals collected from the task was implemented using MATLAB software (The Mathworks Inc.) by using the Fieldtrip Toolbox (Oostenveld et al., 2011). The pre-processing steps were the same as those implemented in the previous lexical decision task. First, an artifact rejection was carried out in order to eliminate trials with amplitude values exceeding $\pm 100 \mu V$. Next, an independent component analysis (ICA) was run to detect and correct visual artifacts, and then a new artifact rejection was applied to ensure the total rejection of artifacts in data. The signal was segmented in periods of 1,500 ms, from -600 to 900 ms post target onset (from -600 to 1,000 ms post target onset for the lexical decision task). The baseline was corrected using the 250 ms preceding the picture/hash mark onset. A new reference was calculated using the mean activity of the mastoid electrodes, applied to 62 electrodes with the activity of the online reference (Cz) recovered. A new sampling rate was established at 256 Hz and a low pass band filter was applied at 30 Hz. ERPs were computed by averaging segments per subject and per condition.

ERP Data Analysis

Visual inspection of ERP waveforms obtained at the present silent reading task revealed a reduction (from first vs. sixth block) in the amplitude of novel word-forms trained under the associative condition, in comparison to those presented under the single orthographic training. Such training effect reached maximum around 300 ms post-stimulus onset at frontal and central scalp electrodes, likely reflecting the different influence of both training conditions in the N400 component (see **Figure 2**). The inspection of the ERP waveforms for the lexicality effect (differences between novel and known word-forms trained in associative condition) also showed a modulation in the N400 latency, for both silent reading and lexical decision tasks (see **Figures 3, 4**). Then, for each task, a temporal window from 285 to 415 ms was selected and the mean activity of known and novel word-forms after the training was extracted in representative midline electrodes (AFZ/3/4, CZ/1/2 and POZ/3/4), where the ERP component of interest (N400) usually peaks at central sites. Two different analyses were carried out to address our hypotheses.



First, we aimed to determine whether the semantic-associative training caused higher facilitation in the lexico-semantic processing of novel word-forms than the non-associative training in a task that was context free of categorization-confounding demands (namely, in the reading task). For this purpose, the effect of the associative and the non-associative conditions was evaluated through the present silent reading task by means of a $2 \times 2 \times 3$ repeated measures ANOVA with training (associative and non-associative), block (first and sixth) and region (frontal, central and posterior) as within-subject factors.

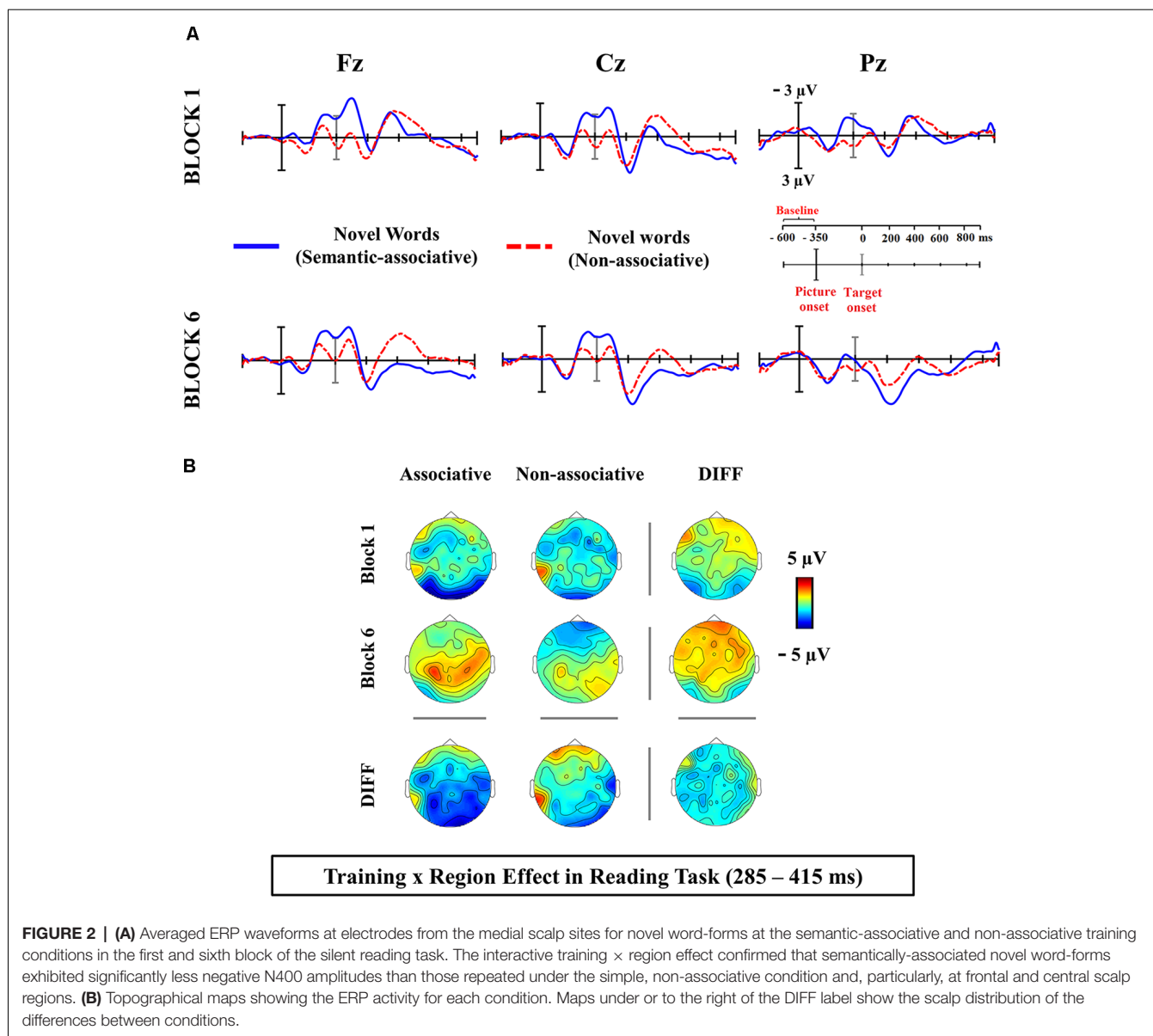
Second, we aimed to further analyze the impact of the associative training in the lexicality effect (namely, in the differences between known and novel word-forms before and after their associative training) in the present reading task as well as at the previous lexical decision task. Thus, a $2 \times 2 \times 3$ repeated measures ANOVA with lexicality (known and novel word-forms), block (first and sixth) and region (frontal, central and

posterior) as within-subject factors, was computed separately for the silent reading and the lexical decision tasks.

RESULTS

Effect of Training in Silent Reading Task

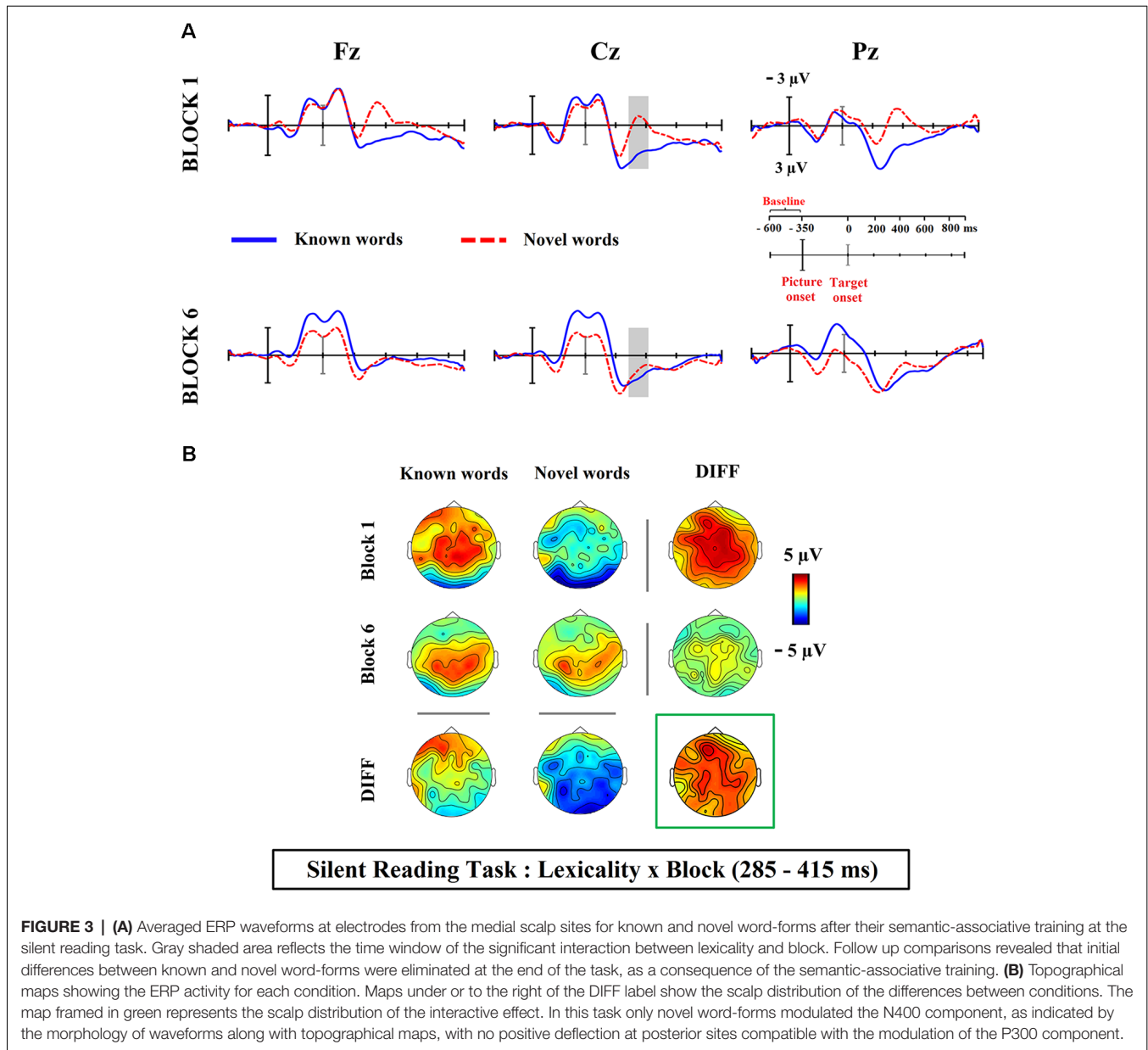
The $2 \times 2 \times 3$ repeated measures ANOVA with the type of training (novel word-forms after associative and non-associative training), block (first and sixth) and region (frontal, central and posterior) conducted for the reading task revealed main effects of block ($F_{(1,24)} = 7.031$, $p = 0.014$, $\eta_p^2 = 0.22$, $1-\beta = 0.72$) and region ($F_{(2,48)} = 4.32$, $p = 0.019$, $\eta_p^2 = 0.15$, $1-\beta = 0.72$), as well as significant interactions between training and region ($F_{(2,48)} = 4.11$, $p = 0.022$, $\eta_p^2 = 0.14$, $1-\beta = 0.70$) and block \times region ($F_{(2,48)} = 5.91$, $\epsilon = 0.77$, $\eta_p^2 = 0.19$, $1-\beta = 0.78$). No other effects or interactions reached significance ($p > 0.05$). The training \times region interaction was tested



again in a separate ANOVA collapsing the two levels of the factor block ($F_{(2,48)} = 4.11$, $p = 0.022$, $\eta_p^2 = 0.14$, $1-\beta = 0.70$). Follow-up comparisons for the effect of training in each scalp region revealed that differences between training conditions were frontally distributed (see topographic maps in **Figure 2**); thus, novel word-forms repeated under the associative training condition exhibited significantly less negative N400 amplitude than those under the non-associative training condition at frontal ($F_{(1,24)} = 9.38$, $p = 0.005$, $\eta_p^2 = 0.28$, $1-\beta = 0.83$; semantic-associative: $-0.69 \mu\text{V}$, non-associative: $-1.91 \mu\text{V}$) and, marginally, at central regions ($F_{(1,24)} = 3.19$, $p = 0.08$, $\eta_p^2 = 0.11$, $1-\beta = 0.40$), but not at posterior scalp sites ($F_{(1,24)} = 0.015$, $p = 0.90$, $\eta_p^2 = 0.001$, $1-\beta = 0.05$; see **Figure 2**). Hence the repeated exposure to novel word-forms under the combination of orthographic and semantic training

resulted in less negative N400 responses than under the simple visual condition, and irrespectively on the task block. Nonetheless, the interaction training \times block, although marginal ($F_{(1,24)} = 3.48$, $p = 0.07$, $\eta_p^2 = 0.12$, $1-\beta = 0.43$), suggests that both training conditions changed differently across blocks, with higher N400 reduction exhibited by semantically associated novel word-forms across blocks (diff.: $-2.23 \mu\text{V}$) than those repeated under the non-associative training condition (diff.: $-0.83 \mu\text{V}$); which in turn increased differences between training conditions, from the first (diff.: $0.01 \mu\text{V}$) to the last task block (diff.: $1.40 \mu\text{V}$).

Therefore, in agreement with previous findings using a lexical decision task as training context, the combination of both orthographic and semantic trainings caused a higher reduction of N400 amplitudes elicited by novel word-forms

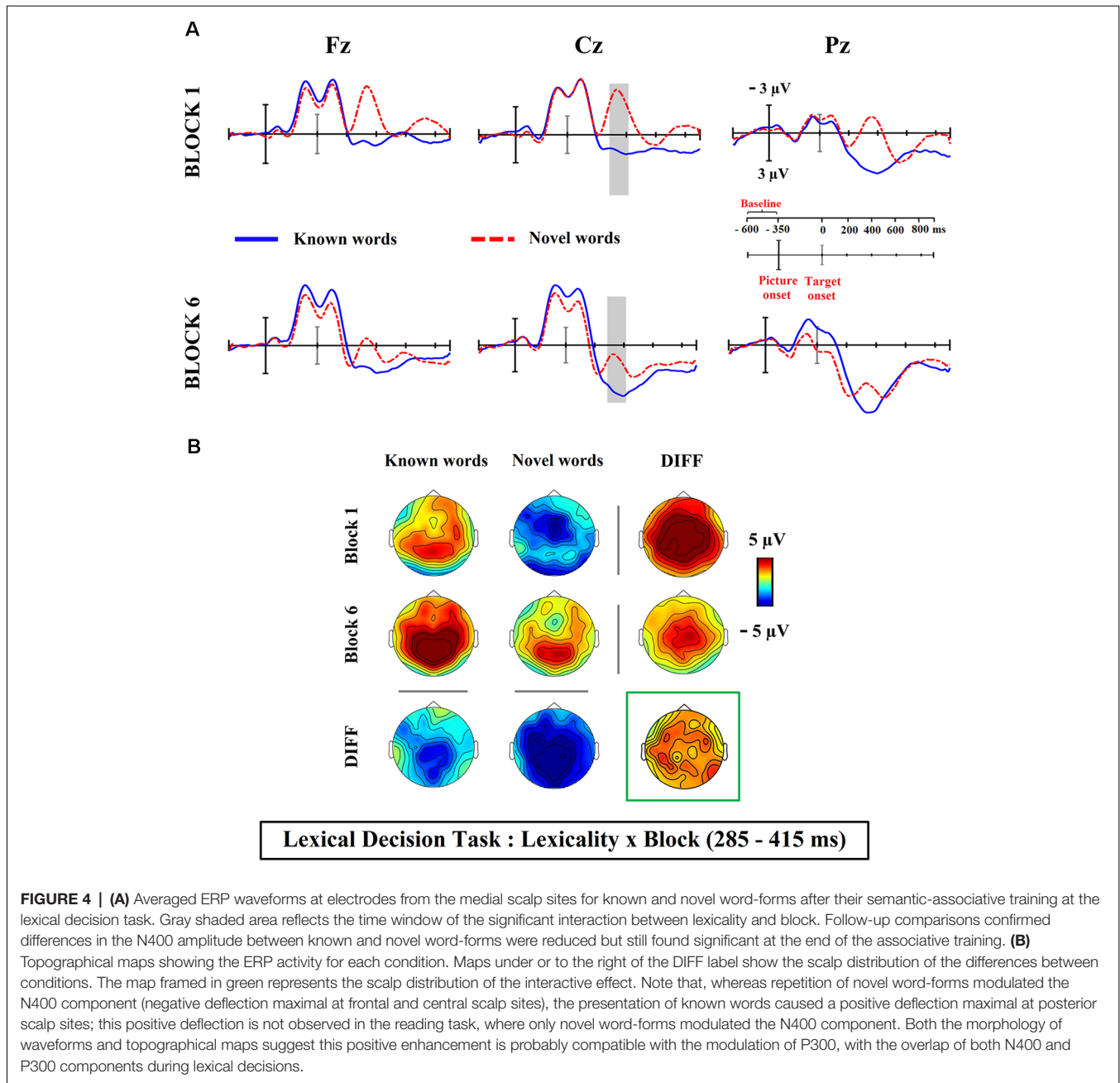


than the simple non-semantic training condition. Importantly, such advantage for the semantic-associative training over the non-associative training was found in the present study in a task free of categorization demands and wherein stimuli are processed automatically.

Changes in Lexicality Effect in the Reading and in the Lexical Decision Tasks

The $2 \times 2 \times 3$ repeated measures ANOVA carried out for the silent reading task with lexicality (known and novel word-forms after semantic-associative training), block (first and sixth) and region (frontal, central and posterior) revealed the main effects of lexicality ($F_{(1,24)} = 10.57$, $p = 0.003$, $\eta_p^2 = 0.30$, $1-\beta = 0.87$) and region ($F_{(2,48)} = 4.86$, $p = 0.012$, $\eta_p^2 = 0.16$, $1-\beta = 0.77$), as

well as lexicality \times block ($F_{(1,24)} = 11.93$, $p = 0.002$, $\eta_p^2 = 0.33$, $1-\beta = 0.91$) and block \times region ($F_{(2,48)} = 8.49$, $p = 0.001$, $\eta_p^2 = 0.26$, $1-\beta = 0.95$) interactions. No other effects were found significant ($p > 0.05$). The lexicality \times block interaction was tested again in a separate ANOVA collapsing the three levels of the factor region ($F_{(1,24)} = 12.03$, $p = 0.002$, $\eta_p^2 = 0.33$, $1-\beta = 0.91$). Follow-up comparisons revealed that differences between novel and known word-forms in the first block ($F_{(1,24)} = 19.64$, $p = 0.000$, $\eta_p^2 = 0.45$, $1-\beta = 0.98$, diff.: 3.45μ V) were eliminated at the end of the training at the sixth block ($F_{(1,24)} = 0.58$, $p = 0.45$, $\eta_p^2 = 0.024$, $1-\beta = 0.11$; diff.: 0.46μ V). Thus, the repetition of novel word-forms at the semantic-associative condition caused a significant modulation in their N400 amplitude across task blocks ($F_{(1,24)} = 9.63$, $p = 0.005$, $\eta_p^2 = 0.28$, $1-\beta = 0.84$, first block:



–1.33 μ V, sixth block: 0.89 μ V), an effect which was not found for known words ($F_{(1,24)} = 0.81$, $p = 0.37$, $\eta_p^2 = 0.033$, $1-\beta = 0.14$, first block: 2.11 μ V, sixth block: 1.36 μ V; see **Figure 3**).

The data set of the lexical decision task (Bermúdez-Margaretto et al., 2015) was submitted to the same $2 \times 2 \times 3$ repeated measures ANOVA, showing significant main effects of lexicity ($F_{(1,21)} = 51.80$, $p = 0.000$, $\eta_p^2 = 0.71$, $1-\beta = 1$) and block ($F_{(1,21)} = 31.39$, $p = 0.000$, $\eta_p^2 = 0.59$, $1-\beta = 1$), as well as lexicity \times block ($F_{(1,21)} = 4.12$, $p = 0.05$, $\eta_p^2 = 0.16$, $1-\beta = 0.49$), lexicity \times region ($F_{(2,42)} = 6.07$, $\epsilon = 0.77$, $\eta_p^2 = 0.22$, $1-\beta = 0.79$) and block \times region ($F_{(2,42)} = 6.37$, $p = 0.004$, $\eta_p^2 = 0.23$, $1-\beta = 0.87$) interactions. The lexicity \times block interaction was

tested again in a separated ANOVA collapsing all three levels of the factor region ($F_{(1,21)} = 4.12$, $p = 0.05$, $\eta_p^2 = 0.16$, $1-\beta = 0.49$). Contrary to results obtained in the silent reading task, follow-up analysis revealed that differences between novel and known words found at the beginning of the semantic-associative training ($F_{(1,21)} = 71.86$, $p = 0.000$, $\eta_p^2 = 0.77$, $1-\beta = 1$; diff.: 4.87 μ V) were reduced but still remained significant at the last block of the training ($F_{(1,21)} = 9.12$, $p = 0.006$, $\eta_p^2 = 0.30$, $1-\beta = 0.82$; diff.: 2.70 μ V, see **Figure 4**). Interestingly, the N400 amplitude resulted as modulated across the lexical decision task not only for novel word-forms ($F_{(1,21)} = 22.67$, $p = 0.000$, $\eta_p^2 = 0.51$, $1-\beta = 0.99$; first block: –2.80 μ V, sixth block: 2.09 μ V) but also

for known words ($F_{(1,21)} = 16.77, p = 0.001, \eta_p^2 = 0.44, 1-\beta = 0.97$; first block: $2.07 \mu V$, sixth block: $4.79 \mu V$). Indeed, known words in the lexical decision task elicited a positive modulation in the N400 time window which was absent in the reading task, as can be observed in the ERP waveforms for known words displayed in **Figure 4**. Such positivity could be compatible with the modulation of the P300 component, with both P300 and N400 components overlapping at the same latency.

Therefore, the semantic-associative training resulted in a different modulation of the N400 lexicality effect at both tasks, with the elimination of differences between known and novel word-forms in the reading task but not in the lexical decision task (although no differences were found between known and novel word forms in reaction times or errors after their semantic-associative training in the lexical decision task, see **Supplementary Material**). However, the positive enhancement elicited by known words in the lexical decision task probably contributed to maintaining lexical differences; indeed, lexicality was found eliminated in the reading task, where such positivity was not enhanced.

DISCUSSION

The present study aimed to determine whether task demands modulate previously reported advantage for novel word lexicalization in the combination of orthographic and semantic-associative training, as compared to single orthographic training. More specifically, we evaluated the impact of these two different training conditions under a more automatic task than that used before for this purpose (lexical decision task), in which an explicit stimuli categorization was required. Thus, we first tested the impact of both types of training on the N400 amplitude through a task free of categorization demands (namely, a silent reading task), in a similar way as previously carried out using a lexical decision task as a training context (Bermúdez-Margaretto et al., 2018). Second, we evaluated the differences in brain activity between newly trained and already known word-forms (i.e., lexicality effect) at both task contexts, as more conclusive proof for the build-up of mental traces into reader's lexicon as a consequence of the training. The results in the present silent reading task confirmed the stronger facilitation in the lexico-semantic processing of novel word-forms after their semantic-associative training in comparison to their single, orthographic training—as reflected in longer reduction of N400 amplitudes for the associative than for the non-associative condition. However, despite the fact that N400 training effect was obtained in both tasks, only those novel word-forms trained in the silent reading task reached a similar lexico-semantic processing to known, already lexicalized words. In contrast, for the lexical decision task, novel words remained showing larger N400 amplitudes than known words after the training, which could be explained by a possible overlap between lexico-semantic and categorization processes, particularly evident for known words. In what follows, ERP findings from both analyses, as well as their implications for novel word learning, are discussed in detail.

The brief exposure to novel written-word forms in association with meaningful cues resulted in the modulation of the

N400 amplitude. Similar findings, indicative of the facilitation in the lexico-semantic processing of novel words, have been reported in prior research after the repetition of these stimuli in association to pictures (Dobel et al., 2010; Angwin et al., 2014; Bermúdez-Margaretto et al., 2018) and definitions (Perfetti et al., 2005; Bakker et al., 2015) or embedding them in meaningful sentence contexts (Mestres-Missé et al., 2007; Borovsky et al., 2010; Frishkoff et al., 2010; Batterink and Neville, 2011). Interestingly, recent research has provided a specific comparison between this meaningful exposition and the single orthographic training of novel words (visual repetition), disentangling both effects and highlighting the advantage of the combined orthographic and semantic training for novel word learning (Bermúdez-Margaretto et al., 2018). In this sense, when both novel word's orthography and meaning were simultaneously trained, higher impact was found in their lexical processing as evidenced in lower N400 amplitudes; in contrast, single orthographic training mainly influenced the episodic processing of these stimuli, as reflected in the LPC enhancement across repeated visual exposures. Nonetheless, a potential confound between lexicalization and categorization processes must be noted in this research, since demands in this task (lexical decisions) could lead to higher discrimination and learning of the stimuli and, importantly, to the acquisition of categorization-guided rather than pure lexical representations for trained word-forms. However, results obtained in the present reading task confirm that, in the absence of such categorization response which could facilitate the learning, a training effect was also obtained, with lower negative N400 amplitudes after combined training. Thus, even in a task without categorization demands and, hence, reflecting likely automatic, and superficial processing of trained stimuli, the passive exposure across meaningful associations leads to a deeper influence in their lexico-semantic instantiation. Therefore, this finding suggests the combination of both orthographic and semantic-associative training could result in a more advantageous strategy for the integration of novel written-word forms into the linguistic system of readers, supporting previous statements. Furthermore, it shows that novel word learning processes can be rather automatic, with the lexico-semantic processing of stimuli accessed and modulated even during a task in which no response is required from readers. Moreover, these findings extend previous results found in this strand of research, which have shown the rapid and automatic acquisition of memory traces for novel written word-forms after their fully unattended, parafoveal exposure (Partanen et al., 2018). In this sense, word learning effects have been found even when reader's attention is directed to different stimuli.

Nonetheless, although such advantage for the associative training is found at both the present silent reading task and at the previous lexical decision task (Bermúdez-Margaretto et al., 2018), the level of automaticity seems to differ between both task contexts, which likely leads to a difference in processing of novel words along their training and hence to their different learning. Indeed, the influence of the task was evident when we evaluated the impact of the meaningful training in the achievement of trained novel words as lexical entities—measured in the N400 lexicality effect. Whereas N400 differences between

novel and known words resulted as being eliminated after the meaningful training in the reading task, the N400 lexicality effect remained significant in the lexical decision task. Such differential N400 lexicality effect is probably a consequence of the influence of categorization processes in this specific task, as particularly evidenced in the brain activity exhibited by known words. In this sense, these stimuli elicited a positive enhancement within the N400 time window; taking into account the positive polarity and more posterior topographical distribution of this effect, the processing of known words being likely to affect the P300 component, with the simultaneous modulation of both N400 and P300 peaks during this task. This component, related to attentional mechanisms activated to accomplish task requirements, such as stimuli categorization (Polich, 1985, 2004; Picton, 1992), is probably reflecting the reader's strategy about the incoming stimuli, addressed to carry out the efficient stimuli categorization during the lexical decision task. Remarkably, such P300 deflection was observable for known but not for novel word-forms, probably contributing to maintaining lexical differences between both stimuli. In this sense, it is possible that the lexico-semantic processing of known words—and consequently the modulation of N400—could be less influenced than for novel words, regardless of their association to meaningful cues. This would lead to a highly evident P300 deflection in the ERP waveforms for known words. In contrast, for novel word-forms, the modulation of the N400 elicited by their repeated association to meaningful cues probably overlaps the activity of the P300 component.

An alternative explanation must also be taken into account; it is also possible that repetition of novel word-forms leads to a lower modulation of P300 for these stimuli than for known words, which were not trained across the task. To further explore this question, future studies should consider the training of known words, as this control could clarify whether lower P300 modulation for trained novel words is a consequence of a decrease in stimuli attention driven by their repeated exposure. Nonetheless, since novel word-forms were required to be categorized, it is rather possible that their lexico-semantic processing was also influenced by the activation of categorization-related processes during learning, as occurred for non-repeated known words, and hence confounding the outcome of the learning. Besides this, other limitations of the present research should be taken into account in future studies by evaluating not only electrophysiological but also behavioral outcomes for the learning of novel words (as well as for general reading abilities in both experimental groups), and testing this process in greater samples than those tested in the tasks reported in this study.

On the other hand, when the task in which the training is carried out does not require a specific categorization response, leading to a more shallow discrimination of the stimuli, no modulation of P300 is observed even in the case of known words. Therefore, the lexico-semantic processing of novel word-forms trained in the reading task was probably not confounded by categorization demands, enabling the construction of mental representations which depend on purely lexical and semantic factors rather than guided by categorization demands. Moreover,

when compared to known words, the lexico-semantic processing of both stimuli is matched, as no categorization was required in this task which could cause the modulation of P300 activity for words, leading to lexical differences between these stimuli and novel words.

Therefore, the present study suggests a probable co-occurrence of both N400 and P300 components in the lexical decision task, reflecting the temporal overlap between different cognitive processes, namely, semantic-associative and task-related, categorization processes. The possible overlap between both components has been discussed in the electrophysiological literature, highlighting that effects attributed to N400 modulations could in fact be caused by an underlying P300 modulation (Rugg, 1990). However, not many studies have empirically explored this ERP co-occurrence in the lexico-semantic domain. For instance, in Roehm et al. (2007), a P300 component was found to be modulated depending on the task, with overlap between this component and N400 when the target was highly predictable and also relevant to solve the tasks. Hence, the P300 deflection observed in the present study is consistent with these previous findings and suggests that the P300 modulation is contributing to the lexicality effect obtained in the lexical decision task.

As claimed in Roehm et al. (2007), effects initially attributed to N400 can actually be influenced by P300 modulations (Bentin, 1987; Kutas and Iragui, 1998; Federmeier and Kutas, 1999). In this regard, the interpretation of the differential lexicality effect found in both tasks should be cautiously addressed, taking into account the simultaneous co-occurrence of both ERP effects. Thus, the remaining N400 lexicality effect is not reflecting the poor lexico-semantic learning of novel word-forms in the lexical decision task; on the contrary, the strong modulation of the N400 amplitude along the task proved the facilitation in the processing of these stimuli. Contrarily, lexical differences are probably maintained as a consequence of the P300 modulation elicited by the categorization of known words. Altogether, these findings suggest that N400 modulations found in language learning paradigms must be carefully explored, considering the possible P300 modulations that can occur simultaneously within the N400 time window as a consequence of task-related strategies. Given the potential confounding between both effects, cautious conclusions about the processes under study must be provided.

In short, the present study confirms the advantage in the processing of novel written words, as a consequence of their semantic-associative repetition, by using a silent reading task free of categorization-confounding demands. Thus, this associative training was found to cause a stronger N400 modulation than the single orthographic exposure even under a low-level demand task, which likely induced the lexicalization of these stimuli as suggested by the elimination of the N400 lexicality effect. Importantly, the brain activity for novel and known word-forms was not found matched when the training was carried out under a lexical decision task. Such differential lexicality effect found across both tasks probably suggests the different influence of each task context in the build-up process of mental representations for novel word-forms: purely related to lexico-semantic processes

in the reading task or possibly confounded by categorization processes in lexical decision. Therefore, this pattern of results indicates the higher suitability of the reading task over the lexical decision task to study the associative learning of novel words in the absence of confounding categorization processes. In this sense, a final remark should be provided regarding the specific task used to address novel word learning. The present study shows that lexico-semantic learning can be effectively studied by using low-level demand tasks, in which no particular response is required from readers. Thus, tasks demanding particular responses and involving higher discrimination of the stimuli, such as lexical decision, should be used with caution to study the lexicalization of novel word-forms, since they introduce processes which are probably not involved during word learning.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

This research was approved by the Ethics Committee of the Psychology Department of the University of Oviedo. Before

starting the experimental tasks, participants received pertinent information about the purpose of the study, the task, and their duration. Then, written informed consent was received from participants.

AUTHOR CONTRIBUTIONS

BB-M conducted the experimental tasks and analyzed the data. BB-M and DB wrote the manuscript. AD and FC designed the experimental tasks.

FUNDING

The article was prepared within the framework of the HSE University Basic Research Program and funded by the Russian Academic Excellence Project '5–100'.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnhum.2019.00347/full#supplementary-material>.

REFERENCES

- Angwin, A. J., Phua, B., and Copland, D. A. (2014). Using semantics to enhance new word learning: an ERP investigation. *Neuropsychologia* 59, 169–178. doi: 10.1016/j.neuropsychologia.2014.05.002
- Bakker, I., Takashima, A., van Hell, J. G., Janzen, G., and McQueen, J. M. (2015). Tracking lexical consolidation with ERPs: lexical and semantic-priming effects on N400 and LPC responses to newly-learned words. *Neuropsychologia* 79, 33–41. doi: 10.1016/j.neuropsychologia.2015.10.020
- Batterink, L., and Neville, H. (2011). Implicit and explicit mechanisms of word learning in a narrative context: an event-related potential study. *J. Cogn. Neurosci.* 23, 3181–3196. doi: 10.1162/jocn_a_00013
- Bentin, S. (1987). Event-related potentials, semantic processes and expectancy factors in word recognition. *Brain Lang.* 31, 308–327. doi: 10.1016/0093-934x(87)90077-0
- Bermúdez-Margaretto, B., Beltrán, D., Cuetos, F., and Domínguez, A. (2018). Brain signatures of new (Pseudo-) words: visual repetition in associative and non-associative contexts. *Front. Hum. Neurosci.* 12:354. doi: 10.3389/fnhum.2018.00354
- Bermúdez-Margaretto, B., Beltrán, D., Domínguez, A., and Cuetos, F. (2015). Repeated exposure to “meaningless” pseudowords modulates LPC, but not N (FN) 400. *Brain Topogr.* 28, 838–851. doi: 10.1007/s10548-014-0403-5
- Borovsky, A., Kutas, M., and Elman, J. (2010). Learning to use words: event-related potentials index single-shot contextual word learning. *Cognition* 116, 289–296. doi: 10.1016/j.cognition.2010.05.004
- Bowers, J. S., Davis, C. J., and Hanley, D. A. (2005). Interfering neighbours: the impact of novel word learning on the identification of visually similar words. *Cognition* 97, B45–B54. doi: 10.1016/j.cognition.2005.02.002
- Carr, T. H., and Dagenbach, D. (1990). Semantic priming and repetition priming from masked words: evidence for a center-surround attentional mechanism in perceptual recognition. *J. Exp. Psychol. Learn. Mem. Cogn.* 16, 341–350. doi: 10.1037/0278-7393.16.2.341
- Coltheart, M., Rastle, K., Perry, C., Langdon, R., and Ziegler, J. (2001). DRC: a dual route cascaded model of visual word recognition and reading aloud. *Psychol. Rev.* 108, 204–256. doi: 10.1037/0033-295x.108.1.204
- Davis, C., and Perea, M. (2005). BuscaPalabras: a program for deriving orthographic and phonological neighborhood statistics and other psycholinguistic indices in Spanish. *Behav. Res. Methods* 37, 665–671. doi: 10.3758/bf03192738
- Deacon, D., Hewitt, S., Yang, C.-M., and Nagata, M. (2000). Event-related potential indices of semantic priming using masked and unmasked words: evidence that the N400 does not reflect a post-lexical process. *Cogn. Brain Res.* 9, 137–146. doi: 10.1016/s0926-6410(99)00050-6
- Dobel, C., Junghöfer, M., Breitenstein, C., Klauke, B., Knecht, S., Pantev, C., et al. (2010). New names for known things: on the association of novel word forms with existing semantic information. *J. Cogn. Neurosci.* 22, 1251–1261. doi: 10.1162/jocn.2009.21297
- Ellis, A. W., Ferreira, R., Cathles-Hagan, P., Holt, K., Jarvis, L., and Barca, L. (2009). Word learning and the cerebral hemispheres: from serial to parallel processing of written words. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 3675–3696. doi: 10.1098/rstb.2009.0187
- Federmeier, K. D., and Kutas, M. (1999). A rose by any other name: long-term memory structure and sentence processing. *J. Mem. Lang.* 41, 469–495. doi: 10.1006/jmla.1999.2660
- Forster, K. I., and Chambers, S. M. (1973). Lexical access and naming time. *J. Verbal Learning Verbal Behav.* 12, 627–635. doi: 10.1016/S0022-5371(73)80042-8
- Frishkoff, G. A., Perfetti, C. A., and Collins-Thompson, K. (2010). Lexical quality in the brain: ERP evidence for robust word learning from context. *Dev. Neuropsychol.* 35, 376–403. doi: 10.1080/87565641.2010.480915
- Glushko, R. J. (1979). The organization and activation of orthographic knowledge in reading aloud. *J. Exp. Psychol. Hum. Percept. Perform.* 5, 674–691. doi: 10.1037/0096-1523.5.4.674
- Jasper, H. H. (1958). The ten twenty electrode system of the international federation. *Electroencephalogr. Clin. Neurophysiol.* 10, 371–375.
- Kiefer, M. (2002). The N400 is modulated by unconsciously perceived masked words: further evidence for an automatic spreading activation account of N400 priming effects. *Cogn. Brain Res.* 13, 27–39. doi: 10.1016/s0926-6410(01)00085-4

- Kiefer, M. (2008). Top-down modulation of unconscious automatic processes: a gating framework. *Adv. Cogn. Psychol.* 3, 289–306. doi: 10.2478/v10053-008-0032-2
- 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). Event-related brain potentials to semantically inappropriate and surprisingly large words. *Biol. Psychol.* 11, 99–116. doi: 10.1016/0301-0511(80)90046-0
- Kutas, M., and Iragui, V. (1998). The N400 in a semantic categorization task across 6 decades. *Electroencephalogr. Clin. Neurophysiol.* 108, 456–471. doi: 10.1016/s0168-5597(98)00023-9
- Kutas, M., and Van Petten, C. (1988). Event-related brain potential studies of language. *Adv. Psychophysiol.* 3, 139–187.
- Kwok, R. K. W., Cuetos, F., Avdyli, R., and Ellis, A. W. (2017). Reading and lexicalization in opaque and transparent orthographies: word naming and word learning in English and Spanish. *Q. J. Exp. Psychol.* 70, 2105–2129. doi: 10.1080/17470218.2016.1223705
- Kwok, R. K. W., and Ellis, A. W. (2015). Visual word learning in skilled readers of English. *Q. J. Exp. Psychol.* 68, 326–349. doi: 10.1080/17470218.2014.944549
- Leach, L., and Samuel, A. G. (2007). Lexical configuration and lexical engagement: when adults learn new words. *Cogn. Psychol.* 55, 306–353. doi: 10.1016/j.cogpsych.2007.01.001
- Maloney, E., Risko, E. F., O'Malley, S., and Besner, D. (2009). Tracking the transition from sublexical to lexical processing: on the creation of orthographic and phonological lexical representations. *Q. J. Exp. Psychol.* 62, 858–867. doi: 10.1080/17470210802578385
- Mestres-Missé, A., Rodriguez-Fornells, A., and Münte, T. F. (2007). Watching the brain during meaning acquisition. *Cereb. Cortex* 17, 1858–1866. doi: 10.1093/cercor/bhl094
- Meyer, M. S., and Felton, R. H. (1999). Repeated reading to enhance fluency: old approaches and new directions. *Ann. Dyslexia* 49, 283–306. doi: 10.1007/s11881-999-0027-8
- Neely, J. H. (2012). “Semantic priming effects in visual word recognition: a selective review of current findings and theories,” in *Basic Processes in Reading*, eds D. Besner and G. W. Humphreys (Hillsdale: Routledge), 272–344.
- 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
- Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J.-M. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG and invasive electrophysiological data. *Comput. Intell. Neurosci.* 2011:156869. doi: 10.1155/2011/156869
- Partanen, E. J., Leminen, A., Cook, C., and Shtyrov, Y. (2018). Formation of neocortical memory circuits for unattended written word forms: neuromagnetic evidence. *Sci. Rep.* 8:15829. doi: 10.1038/s41598-018-34029-y
- Perfetti, C. A. (1985). *Reading ability*. Oxford University Press, New York.
- Perfetti, C. A., Wlotko, E. W., and Hart, L. A. (2005). Word learning and individual differences in word learning reflected in event-related potentials. *J. Exp. Psychol. Learn. Mem. Cogn.* 31, 1281–1292. doi: 10.1037/0278-7393.31.6.1281
- Picton, T. W. (1992). The P300 wave of the human event-related potential. *J. Clin. Neurophysiol.* 9, 456–479. doi: 10.1097/00004691-199210000-00002
- Polich, J. (1985). Semantic categorization and event-related potentials. *Brain Lang.* 26, 304–321. doi: 10.1016/0093-934x(85)90045-8
- Polich, J. (2004). Clinical application of the P300 event-related brain potential. *Phys. Med. Rehabil. Clin. N. Am.* 15, 133–161. doi: 10.1016/s1047-9651(03)00109-8
- Qiao, X., and Forster, K. I. (2013). Novel word lexicalization and the prime lexicality effect. *J. Exp. Psychol. Learn. Mem. Cogn.* 39, 1064–1074. doi: 10.1037/a0030528
- Qiao, X., Forster, K., and Witzel, N. (2009). Is banara really a word? *Cognition* 113, 254–257. doi: 10.1016/j.cognition.2009.08.006
- Roehm, D., Bornkessel-Schlesewsky, I., Rösler, F., and Schlewsky, M. (2007). To predict or not to predict: influences of task and strategy on the processing of semantic relations. *J. Cogn. Neurosci.* 19, 1259–1274. doi: 10.1162/jocn.2007.19.8.1259
- Rugg, M. D. (1990). Event-related brain potentials dissociate repetition effects of high- and low-frequency words. *Mem. Cognit.* 18, 367–379. doi: 10.3758/bf03197126
- Rugg, M. D., and Curran, T. (2007). Event-related potentials and recognition memory. *Trends Cogn. Sci.* 11, 251–257. doi: 10.1016/j.tics.2007.04.004
- Schneider, W., Eschman, A., and Zuccolotto, A. (2002). *E-Prime (Version 2.0). Computer Software and Manual*. Pittsburgh, PA: Psychology Software Tools Inc.
- Share, D. L. (1995). Phonological recoding and self-teaching: sine qua non of reading acquisition. *Cognition* 55, 151–218. doi: 10.1016/0010-0277(94)00645-2
- Share, D. L. (2008). Orthographic learning, phonological recoding and self-teaching. *Adv. Child Dev. Behav.* 36, 31–82. doi: 10.1016/s0065-2407(08)00002-5
- Snodgrass, J. G., and Vanderwart, M. (1980). A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity and visual complexity. *J. Exp. Psychol. Hum. Learn.* 6, 174–215. doi: 10.1037/0278-7393.6.2.174
- Suárez-Coalla, P., Álvarez-Cañizo, M., and Cuetos, F. (2016). Orthographic learning in Spanish children. *J. Res. Read.* 39, 292–311. doi: 10.1111/1467-9817.12043
- Walczyk, J. J. (2000). The interplay between automatic and control processes in reading. *Read. Res. Q.* 35, 554–566. doi: 10.1598/rrq.35.4.7

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer MV declared a shared affiliation, though no other collaboration, with several of the authors (DB, AD) to the handling Editor.

Copyright © 2019 Bermúdez-Margaretto, Beltrán, Cuetos and Domínguez. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Language Processing as a Precursor to Language Change: Evidence From Icelandic

Ina Bornkessel-Schlesewsky^{1,2*}, Dietmar Roehm^{3,4}, Robert Mailhammer^{5,6} and Matthias Schlesewsky^{1,2}

¹ Cognitive and Systems Neuroscience Research Hub, University of South Australia, Adelaide, SA, Australia, ² School of Psychology, Social Work and Social Policy, University of South Australia, Adelaide, SA, Australia, ³ Centre for Cognitive Neuroscience, University of Salzburg, Salzburg, Austria, ⁴ Department of Linguistics, University of Salzburg, Salzburg, Austria, ⁵ School of Humanities and Communication Arts, Western Sydney University, Penrith, NSW, Australia, ⁶ The MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Penrith, NSW, Australia

OPEN ACCESS

Edited by:

Olga V. Shcherbakova,
Saint Petersburg State
University, Russia

Reviewed by:

Zude Zhu,
Jiangsu Normal University, China
Paolo Canal,
University Institute of Higher Studies in
Pavia, Italy

*Correspondence:

Ina Bornkessel-Schlesewsky
Ina.Bornkessel-Schlesewsky@
unisa.edu.au

Specialty section:

This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Psychology

Received: 19 April 2019

Accepted: 19 December 2019

Published: 17 January 2020

Citation:

Bornkessel-Schlesewsky I, Roehm D,
Mailhammer R and Schlesewsky M
(2020) Language Processing as a
Precursor to Language Change:
Evidence From Icelandic.
Front. Psychol. 10:3013.
doi: 10.3389/fpsyg.2019.03013

One of the main characteristics of human languages is that they are subject to fundamental changes over time. However, because of the long transitional periods involved, the internal dynamics of such changes are typically inaccessible. Here, we present a new approach to examining language change via its connection to language comprehension. By means of an EEG experiment on Icelandic, a prominent current example of a language in transition, we show that the neurophysiological responses of native speakers already reflect projected changes that are not yet apparent in their overt behavior. Neurocognitive measures thus offer a means of predicting, rather than only retracing, language change.

Keywords: language comprehension, language change, event-related potentials, Icelandic, N400, late positivity

1. INTRODUCTION

Since the earliest days of the human species, human culture, and society have undergone a continuous series of changes and adaptations. Language, as the primary means of human communication, has always played an integral role in this process. English is a particularly good example of how profound such changes can be. From Old English (~400–1100 AD) to Modern English, the language has undergone at least two radical transitions: word order became fixed and the language's rich morphological system (e.g., case inflections) was drastically reduced. The communicative consequences of these changes are profound, because the properties in question crucially influence the way in which meaning can be extracted from the speech stream in real time (Bornkessel-Schlesewsky et al., 2015). In modern English, the fixed positioning of elements allows hearers to determine “who is doing what to whom” in a strictly linear manner (in transitive sentences with default verb classes, the Actor performing the action precedes the Undergoer affected by that action). In Old English or other Germanic languages such as modern German, by contrast, these relations are determined less by linear position and rather mainly by the form in which the event participants are expressed (e.g., via nominative or accusative case marking).

Between the thirteenth and fifteenth centuries (Allen, 1995), English underwent a transition from a grammar favoring morphological marking as the primary means of expressing participant roles in a sentence (“Grammar A”), to a grammar using linear position to the same purpose (“Grammar B”). The tendency toward such a change is a key property of the Germanic language family as a whole (Faarlund, 2001; Platzack, 2002).

However, the internal dynamics of the transitional processes from Grammar A to Grammar B appear virtually inaccessible, since an observation of the relevant changes (e.g., in spoken or written language) presupposes that they have already taken place in individuals. And even zooming in on the individual, the question is what exactly triggers changes in speech behavior? Theories on language change offer different perspectives on what causes change and how it begins¹. Perception-based approaches generally assume that the perception or interpretation of input material changes in individuals, and that this is then transferred onto production. Examples for this type of perspective are speech-perception-based models of sound change (e.g., Ohala, 1981), and mainstream models of grammaticalization (e.g., Hopper and Traugott, 1993). By contrast, production-based models see changes as by-products of production (see e.g., Bybee, 2010; Harrington, 2012). For the types of changes mentioned above, i.e., changes in the way in which grammatical relations and semantic roles are indicated, both approaches are relevant, albeit to different degrees. The loss of morphological marking is a typical consequence of phonological erosion, a result of production (Bybee, 2010). In the case of English, for example, the phonological reduction and loss of unstressed final syllables is typically seen as a consequence of a fixed dynamic accent at the left edge of words. The utilization of constituent order to express grammatical relations, however, is best understood as an effect of processing information that is increasingly ambiguously marked. In fact, ambiguous structures are usually seen as the key element for triggering a reinterpretation in most theories of grammaticalization. This, in turn, leads to the recruitment of constituent ordering for purposes of expressing grammatical relations.

From the perspective of perception-based approaches to language change, then, reinterpretation by hearers precedes overt changes in how language is produced by speakers. Accordingly, overt manifestations of language change in speech or writing should be preceded by measurable preparatory changes in neural language comprehension mechanisms. We propose that this hypothesis may be investigated by measuring the brain activity of individuals who speak a language in transition. Electrophysiological measures appear particularly well-suited to revealing such effects as they (a) allow us to observe distinctions that are not consciously accessible to speakers (e.g., Bornkessel et al., 2004), and (b) have been shown to be sensitive to the transitional phenomena under examination here, namely word order and case marking (for reviews, see Bornkessel and Schlewsky, 2006a; Bornkessel-Schlesewsky and Schlewsky, in press). From this perspective, the relation between language processing and language change constitutes an intriguing challenge for examining the brain-behavior interface. If our assumptions are correct, it may eventually be possible to

¹Note that we do not consider sociolinguistic perspectives on language change here, as they are outside the scope of what we aim to examine with this paper. Our study investigates whether shifts in processing (i.e., comprehension) strategies in individuals precede overt language change, i.e., if speakers change their perception before the propagation of any change that would require a sociolinguistic perspective, and if this could possibly be a trigger for language change.

link processing phenomena at the timescale of several hundred milliseconds to cross-generational changes in language use (cf. also Christiansen et al., 2016).

Here, we take a first step toward testing the hypothesis that changes in language comprehension may precede overt language change by comparing electrophysiological correlates of sentence comprehension to judgements of sentence acceptability in Icelandic. Within the Germanic language family, Icelandic stands out for its parallels to English during the transitional period. Thus, it has a fully fledged system of morphological case marking including non-nominative subjects, but shows considerable word order strictness such as a fixed subject position (Zaenen et al., 1990; Thrainsson, 2014)². In terms of linear order, Icelandic therefore behaves very similarly to Modern English (Grammar B), while its morphological properties render it more closely comparable to earlier stages of the English language (Grammar A). In addition, there are initial indications that the morphological (case) system is becoming unstable, as speakers are showing an increasing tendency to reduce the number of different case forms that can occur in particular linear positions in the sentence—a phenomenon known as “case sickness” (Smith, 1994; Eythórsson, 2000). As discussed in detail by Smith (1994), two alternations of this type “occur in most Germanic languages at some stage” (Smith, 1994, p. 675): the tendency for accusative subjects of experiencer verbs to be marked with dative (dative substitution, DS) and the tendency for accusative or dative subject marking to be replaced by nominative (nominative substitution, NS). The following Icelandic examples from Jónsson and Eythórsson (2005) illustrate DS (example 1) and NS (example 2), respectively (see their paper for further examples from Faroese and Smith, 1994 for examples from other Germanic languages such as German and Old English):

- (1) a. **Mig** vantar hnif.
me-ACC need-3.SG knife
b. **Mér** vantar hnif.
me-DAT need-3.SG knife
“I need a knife.”
- (2) a. **Bátnum** hvolfði á flóanum.
boat.the-DAT capsized-3.SG on bay.the
b. **Báturinn** hvolfði á flóanum.
boat.the-NOM capsized-3.SG on bay.the
“The boat capsized on the bay.”

Nominative substitution parallels the diachronic changes that took place in the history of English, as a result of which the dative or accusative marking of experiencer arguments was replaced with nominative. This is illustrated by the examples in 3 (cited from Smith, 1994) using the verb *ofhreowan* (“to

²From the perspective of traditional grammar, “non-nominative subject” may appear to be a contradiction in terms because nominative marking is one of the classic defining properties of subjecthood. However, cross-linguistic research has revealed the existence of non-nominative subjects in a number of languages in all parts of the world (e.g., Hindi, Russian, Japanese) (Bhaskararao and Subbarao, 2004). While these arguments do not bear nominative case and often show no (or only reduced) agreement with the verb, they display a number of other subject properties e.g., with respect to control, reflexivization, deletion under conditions of coreference.

pity”). Both example sentences stem from the writings of Ælfric of Eynsham, an English abbot who lived in the tenth and eleventh centuries AD.

- (3) a. him ofhreow þæs mannes
him-DAT was-pity the-GEN man-GEN
“He was sorry for the man.” (*ÆCHom* I 13 192.16)
- b. se mæssepreost þæs mannes ofhreow
that-NOM priest that-GEN man-GEN pitied
“The priest took pity on the man.” (*ÆLS* (Oswald) 262)

1.1. The Present Study

The present study used event-related potentials (ERPs) to investigate how native speakers of Icelandic process constructions that differ in regard to their compatibility with the target grammar of change (Grammar B). While we did not contrast language comprehension and production directly, we compared electrophysiological correlates of language comprehension to participants’ own acceptability judgements as a first step toward a full-fledged examination of the perception-driven hypothesis of language change. Thus, we compared participants’ neural responses to their overt, language-related behavior. As already discussed above, ERP responses do not directly reflect individuals’ conscious assessment of sentence wellformedness (e.g., Bornkessel et al., 2003, 2004; Bornkessel and Schlewsky, 2006b), rather mirroring the demands of online sentence processing (see also Demiral et al., 2008). This allows us to examine potential shifts in these demands vis-à-vis how participants view their own language. If overt language change has already taken place, new structures will be both produced by speakers and judged to be acceptable by hearers, even though they may not be considered grammatical from a prescriptive perspective. Here, we examine the extent to which neural language processing and acceptability judgements differ from (prescriptive) grammatical assumptions and whether this comparison can shed new light on the dynamics of language change.

Participants were presented with two critical types of sentences (see **Table 1**): structures with an initial dative and a post-verbal nominative and structures with an initial nominative and a post-verbal dative. We assume that the nominative-before-dative sequence is the target structure on which a fully completed transition to (the Modern English type) Grammar B will finally converge. We can thus use the differential brain response to these structures as opposed to their dative-before-nominative counterparts as a diagnostic tool for how far the neural transition toward Grammar B has advanced.

In order to characterize the degree of transition more closely, we used three verb types which differ in their compatibility between Grammar A and Grammar B: (a) active verbs, which were already associated with a nominative-before-dative structure in Grammar A and thus do not require a change to be compatible with Grammar B; (b) dative subject-experiencer verbs, which are obligatorily associated with a dative-before-nominative structure in Grammar A (and current Icelandic) and must thus undergo a transition to nominative-before-dative to be compatible with Grammar B; and (c) alternating verbs, which

TABLE 1 | Example sentences for the present study.

Verb	NP2 case	Example
ACT	DAT	... drekkt / fisksalanum / í brunninum. ... drowned fish-salesman-DAT in well-the
	NOM	*... drekkt / fisksalinn / í brunninum. ... drowned fish-salesman-NOM in well-the
ALT	DAT	... fylgt / konunni / í borginni. ... followed lady-DAT in city-the
	NOM	... fylgt / konan / í borginni. ... followed lady-NOM in city-the
EXP	DAT	*... mislíkað / nemandanum / á kaffihúsi. ... disliked student-DAT in coffeehouse-the
	NOM	... mislíkað / nemandinn / á kaffihúsi. ... disliked student-NOM in coffeehouse-the

Note that all sentences commenced with a main clause that was common to each item, e.g., *Ég vantroysti sjómanninum / sem / hefur ...*, ‘I distrust seaman-the-DAT who has...’, ‘I distrust the seaman who has ...’ The critical position (NP2 in the relative clause) is marked in bold and segmentation for visual presentation is indicated by the forward slashes (/).

are already in transition between Grammar A and Grammar B in that they allow both a nominative-before-dative and a dative-before-nominative order (Barðdal, 2001).

Note that, strictly speaking, the different verb types used here are in fact associated with changes in subject case rather than just word order. Dative subject-experiencer verbs require a dative subject and nominative object. Alternating verbs, by contrast, are compatible with a nominative subject and dative object as well as with a dative subject and nominative object. Finally, active verbs require a nominative subject and dative object.

The relative clause constructions used here served to create a fixed subject-before-object word order. The subject is expressed by the relative pronoun *sem*, which is coreferent with the noun in the main clause. As *sem* is invariant across different cases, it does not become clear until the post-verbal noun in the relative clause (NP2) whether the word order is nominative-before-dative or dative-before-nominative. As the verb has already been processed at this point, NP2 is the critical position for observing expectation mismatches in regard to the word order/case marking. Based on previous ERP experiments that examined case marking and word order in several languages including German, Swedish, Japanese, and Hindi, we expect such mismatches to be reflected in an N400 followed by a late positivity (e.g., Frisch and Schlewsky, 2001; Bornkessel et al., 2004; Mueller et al., 2005; Haupt et al., 2008; Choudhary et al., 2009; Hörberg et al., 2013). We will return to our proposed functional interpretation of these components—and how this may relate to language change—in the discussion section.

2. METHODS

2.1. Participants

Twenty-three students from the University of Iceland (Reykjavik) participated in the experiment [13 female, mean age 25.39 (sd = 3.71) years, age range 17–30 years]. All participants were right-handed native speakers of Icelandic with normal or

corrected-to-normal vision and gave written informed consent before the experimental session. Seven additional participants were excluded from the final data analysis due to varying numbers of trials per condition and a different task setup (acceptability task only): these were the first seven participants run, on the basis of which we concluded that the experimental protocol was too long and that a second task was required in order to avoid strategic effects.

2.2. Materials

Each sentence consisted of a matrix clause with a first person nominative-subject (*ég* “I”) and one of four nominative-subject experienter verbs (*vantreyisti* “distrust”; *treysti* “trust”; *man eftir* “remember”; *trúi* “believe”), which were distributed equally across conditions and were followed by a dative case marked noun (object) and a subsequent subject relative clause relating to it. The subject relative clause began with the inflexible relative pronoun *sem*, which is fully case ambiguous (NOM/DAT/GEN/ACC). Note that, in contrast to English, the relative pronoun *sem* must always be at the beginning of a clause and can never be preceded by a preposition (e.g., *húsið, sem hann bjó í* “The house that he lived in”). In addition, the relative pronoun cannot be dropped from the clause as in English (a general restriction in other Germanic languages). A finite auxiliary and the main verb followed the relative pronoun, thereby explicitly indicating that *sem* refers to the subject of the relative clause. After the main verb of the relative clause, there was a case-marked noun followed either by a temporal, local, reason, or manner adverbial. The type of verb within the relative clause was manipulated according to the design in **Table 1**: active verbs, alternating verbs, and dative-subject experienter verbs (the choice of verbs was motivated by Barðdal, 2001; Jónsson, 2003). For each verb type, two different sentence types were created: the post-verbal noun was either marked with dative or nominative case marking. Participants read 48 sentences in each of the two conditions with active verbs, 24 sentences in each of the two conditions with alternating verbs, and 20 sentences in each of the conditions with dative-subject experienter verbs, thus resulting in a total of 184 sentences. The differing trial numbers between verb classes were chosen so as to ensure an equal split between default (active) and non-default (alternating, dative-subject experienter) verbs. Sentences were presented to participants in a pseudo-randomized manner.

2.3. Procedure

Sentences were presented visually in the center of a computer screen. Each trial began with the presentation of an asterisk (1,000 ms) in order to fixate participants’ eyes at the center of the screen and to alert them to the upcoming presentation of the sentence. Main clauses were presented as a single chunk (1,000 ms), followed by a word-by-word presentation of the relative clause. Each word was presented for 750 ms (adverbials were presented as chunks), with an inter-stimulus interval (ISI) of 150 ms. This relatively long presentation time was chosen because of the morphological complexity of the language (for similar arguments for Turkish, see Demiral et al., 2008) and was perceived as a comfortable reading rate by participants. After the

presentation of the sentence, there was a 400 ms pause before participants were required to complete an acceptability judgment task (signaled through the presentation of a question mark), which involved judging whether the sentence was acceptable or not. Participants responded by pressing the left or right mouse button for “yes” or “no.” The time window for the button press was restricted to 3,000 ms. Afterwards, participants responded to a comprehension question (an indirect interrogative sentence querying actor/undergoer roles). Again, the maximal reaction time for this task was 3,000 ms. Trials were separated by an inter-trial interval (ITI) of 1,250 ms.

Participants were asked to avoid movements and eye-blinks during the presentation of the sentences. All experimental sessions began with a short training session followed by 8 experimental blocks, between which the participants took short breaks. Each experimental session lasted ~2 h (including electrode preparation).

2.4. EEG Recording and Preprocessing

The EEG was recorded by means of 29 sintered Ag/AgCl-electrodes fixed at the scalp by means of an elastic cap (Easy Cap, Herrsching-Breitbrunn, Germany). The ground electrode was positioned at C2. Recordings were referenced to the left mastoid. The electrooculogram (EOG) was monitored by means of electrodes placed at the outer canthus of each eye for the horizontal EOG and above and below the participant’s left eye for the vertical EOG. Electrode impedances were kept below 5 kOhm. All EEG and EOG channels were amplified using a BrainVision BrainAmp amplifier (time constant 10 s, high cutoff 250 Hz) and recorded with a digitization rate of 500 Hz.

EEG data were preprocessed using MNE Python version 0.19.1 (Gramfort et al., 2013, 2014) supplemented by additional utility functions from the philistine package (<https://gitlab.com/palday/philistine>). EOG artifacts were corrected using Independent Component Analysis (ICA). To this end, a copy of the raw data was bandpass filtered from 1 to 40 Hz (zero-phase, hamming windowed FIR filter; length: 1,651 samples; transition bandwidth: 1–10 Hz). ICAs were computed using the FastICA method with 25 components (EEG channels only; epochs with peak-to-peak voltages exceeding 250 microvolts were excluded from consideration). We used the “create_eog_epochs” function in MNE to find EOG events; these were then used to identify EOG-related ICs via correlation (function “ica.find_bads_eog”). The components thus identified were removed from the original raw data. Subsequently, the data were filtered with a 0.1–30 Hz bandpass filter (zero-phase, hamming windowed FIR filter; filter length: 16,501 samples; transition bandwidth: 0.1–7.5 Hz) to exclude slow signal drifts and high frequency noise. The data were epoched from –200 to 1,200 ms relative to the onset of the critical second NP. Epochs with peak-to-peak amplitudes exceeding 150 microvolts for EEG channels were excluded, as were flatlining epochs with peak-to-peak voltages under 5 microvolts. No baseline correction was applied; rather, the trial-by-trial mean prestimulus voltage (–200 to 0 ms) was included as a covariate in the statistical analysis and used to baseline-correct the plots (Alday, 2019).

2.5. Data Analysis

We used R Version 3.6.1 for all statistical analyses (R Core Team, 2018) and the packages tidyverse version 1.2.1 (Wickham et al., 2019), lme4 version 1.1.21 (Bates et al., 2015), car version 3.0-4 (Fox and Weisberg, 2011), emmeans version 1.4.2 (Lenth, 2019), and cowplot version 1.0.0 (Wilke, 2019). Raincloud plots were produced using the method and code supplied by Allen et al. (2019). To produce model output tables, we used lmerOut version 0.5 (Alday, 2018) and kableExtra version 1.1.0 (Zhu, 2019). Raw data and all analysis scripts are available via the Open Science Framework (see Data Availability Statement.)

For all analyses below, contrasts for categorical factors used sum coding (for a tutorial on contrast coding, see Schad et al., 2020), i.e., coefficients reflect differences to the grand mean.

2.5.1. Behavioral Data

Behavioral data were analyzed using generalized mixed effects models with fixed effects verb and case and random intercepts by participant and item. More complex random effect structures involving random slopes by participant and item did not converge.

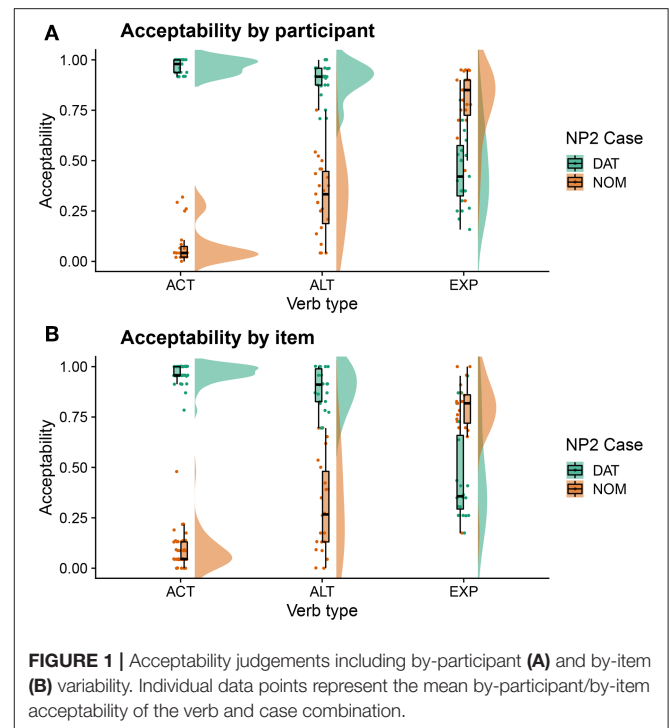
2.5.2. EEG Data

Single-trial EEG data were analyzed using mixed effects models with fixed effects verb, case, and epoch (i.e., time within the experiment), topographical factors laterality and sagittality and their interaction. Laterality and sagittality were implemented as continuous predictors so as to provide a more fine-grained perspective on topographical similarities and differences between individual electrodes (see Brilmayer et al., 2019). To this end, we used positional coordinates retrieved from http://robertoostenveld.nl/electrodes/besa_81.txt. We standardly include epoch as a fixed effect when analysing EEG data in order to examine whether effects change over the course of the experiment. Individual trial mean prestimulus EEG amplitude (−200 to 0 ms) was included in the model as a covariate in lieu of baseline correction (Alday, 2019). (See also Alday and Kretzschmar, 2019, for an example of this approach). Epoch and prestimulus EEG amplitude were centered prior to their inclusion in each model. Models also included random slopes for the interaction of verb and case by participant and for case by item. More complex random effects structures including trial led to convergence problems. We analyzed single-trial ERP amplitudes in the following two time windows: 300–500 ms for the N400 and 700–1,000 ms for the late positivity.

3. RESULTS

3.1. Behavioral Data

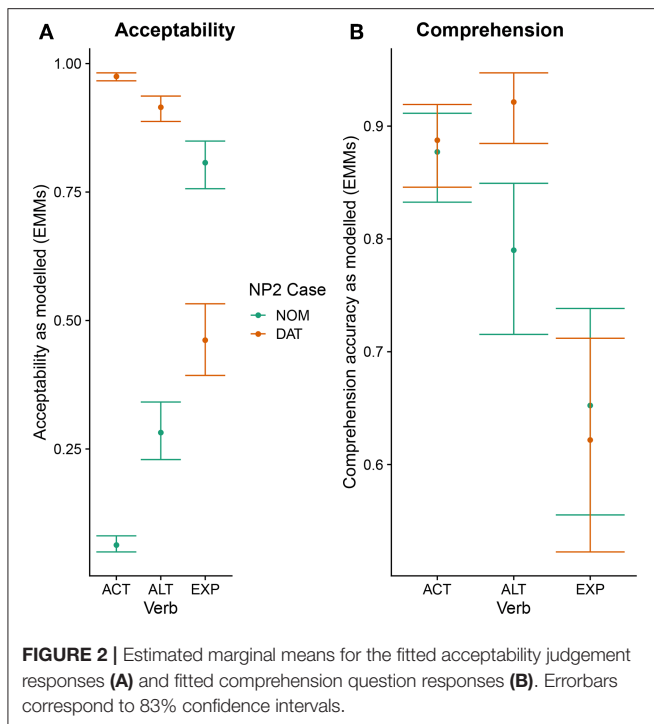
The results of the acceptability judgement task are visualized in **Figure 1** using raincloud plots (Allen et al., 2019). **Figure 1A** shows variability by participant, i.e., individual data points represent the mean by-participant acceptability of each verb and case combination. **Figure 1B**, by contrast, shows variability by item, i.e., individual data points represent the mean by-item acceptability of each verb and case combination. As is apparent from the figure, active verbs showed a clear preference



for a dative-marked NP2, i.e., for nominative-dative orders. This was the case both by participants and items. Alternating verbs also showed a general preference for nominative-dative orders, but with a less clear-cut pattern than active verbs. While nominative-dative orders were highly acceptable for all participants and items, there was considerably more variability for dative-nominative orders. Finally, experiencer verbs showed an overall preference for dative-nominative orders. However, there was again considerable variability underlying this pattern. Participants varied widely with regard to how acceptable they found both orders, i.e., some participants accepted the—supposedly ungrammatical—nominative-dative order and some tended to reject the dative-nominative order. A similar pattern emerged by item.

Statistical analysis of the acceptability data using generalized linear mixed effects modeling revealed main effects of verb [type II Wald test: $\chi^2(2) = 13.97$, $p < 0.001$] and case [$\chi^2(1) = 245.57$, $p < 0.001$], as well as an interaction between the two [$\chi^2(2) = 856.31$, $p < 0.001$]. Model estimates are visualized in **Figure 2A** using estimated marginal means. This also serves to resolve the interaction. The errorbars in this and the following figures represent 83% confidence intervals, the non-overlap of which corresponds to significance at the 5% level.

For the comprehension task, participants had a mean accuracy of 75% (sd: 22%). Generalized linear mixed effects modeling again showed main effects of verb [type II Wald test: $\chi^2(2) = 46.51$, $p < 0.001$] and case [$\chi^2(1) = 9.16$, $p < 0.01$], as well as an interaction between the two [$\chi^2(2) = 31.37$, $p < 0.001$]. Model estimates are visualized in **Figure 2B** using estimated marginal means. As is apparent from the figure, participants showed a high comprehension accuracy for both word orders with active verbs.



In sentences with alternating verbs, by contrast, comprehension was significantly more accurate for nominative-dative than for dative-nominative orders. Finally, for the experiencer verbs, comprehension accuracy was relatively low for both word orders.

Full model summaries for the behavioral data are presented in Tables S1, S2.

3.2. ERP Data

Grand average ERPs at the critical NP2 position within the relative clause are shown in Figures 3–5 for active, alternating and experiencer verbs, respectively. Active verbs show a biphasic N400–late positivity pattern for dative-nominative vs. nominative-dative orders (i.e., for sentences in which NP2 is marked nominative as opposed to dative). A similar but less pronounced pattern is observable for the alternating verbs. Experiencer verbs, by contrast, show a slight tendency for a reversed pattern in the N400 (i.e., increased negativity for nominative-dative vs. dative-nominative orders), but there is no indication of differences in the late positivity. By-participant and by-item variability in the ERPs are visualized in Figures S1–S6. These show that variability by both participants and items is higher for alternating and experiencer verbs in comparison to active verbs.

The ERP data were analyzed using linear mixed effects models as outlined above. Tables 2, 3 provide a broad summary of effects in the N400 and late positivity time windows, respectively, using Type II Wald tests. Full model summaries are presented in Tables S3–S6. In line with our hypotheses, we focus on interactions of verb type and case and, for each statistical model, interpret the highest-order interaction involving both of these factors.

In the N400 time window, Wald tests revealed an interaction of verb \times case \times sagittality \times epoch. This interaction is resolved and visualized in Figure 6, which shows estimated marginal means and 83% confidence intervals. As noted above for the behavioral data, non-overlap of 83% confidence intervals corresponds to a significant difference at the 5% level. It is apparent from Figure 6 that, for active verbs, dative-nominative orders show a negativity in comparison to nominative-dative orders over the course of the entire experiment. This effect is clearest in central and posterior regions. Alternating verbs, by contrast, do not show a clear pattern at the beginning of the experiment, but an N400 effect for dative-nominative vs. nominative-dative orders emerges over time and is clearly apparent in central and posterior regions by the end of the experiment. Experiencer verbs do not show any differential N400 effects for the two word orders at any point over the course of the experiment. Figures S7–S9, which serve to resolve the additional prestimulus interval \times verb \times case \times sagittality \times epoch interaction, show that this overall pattern is broadly consistent across a range of values of prestimulus amplitude from -5 to $5 \mu V^3$.

Please note that the relatively broad distribution of the N400 effects observed here (i.e., the fact that these effects weren't confined to centro-posterior sites but were also observable at more anterior channels) is consistent with the existing literature. A number of previous studies examining case-based processing mismatches have reported similarly broad N400 distributions (e.g., Frisch and Schlesewsky, 2001; Mueller et al., 2005).

For the late positivity time window, Wald tests showed an interaction of prestimulus amplitude \times verb \times case \times epoch, which is resolved and visualized in Figure 7. Active verbs show a clear positivity for dative-nominative vs. nominative-dative sentences. For alternating verbs, a similar effect emerges over the course of the experiment. Finally, experiencer verbs show no indication of a late positivity effect for one word order as compared to the other.

Figures S10–S12 illustrate the verb \times case \times epoch interaction for different values of prestimulus amplitude. As for the N400, effects are consistent across a range of prestimulus values.

3.3. Acceptability-Contingent Analyses of ERPs to Dative Subject-Experiencer Verbs

For the dative subject-experiencer verbs, we conducted an additional analysis in order to examine whether the overall component pattern—i.e., the absence of N400 / late positivity effects differentiating between word orders—might be a reflection of the high variability of acceptability ratings for these verbs (cf. Figure 1). To this end, we fit a mixed model to the experiencer verb data in which we added single trial acceptability

³As Alday (2019) notes in regard to interpreting interactions with prestimulus amplitudes: “As elsewhere in statistics, we can include additional covariates as controls without further interpreting those covariates. In other words, we can safely ignore the terms related to baseline correction, but we cannot omit them from the model.” (p. 9) We include the supplementary figures for completeness' sake and to explore whether the presence of different verb classes prior to our critical NP2 position may have had an effect. However, this does not appear to have been the case.

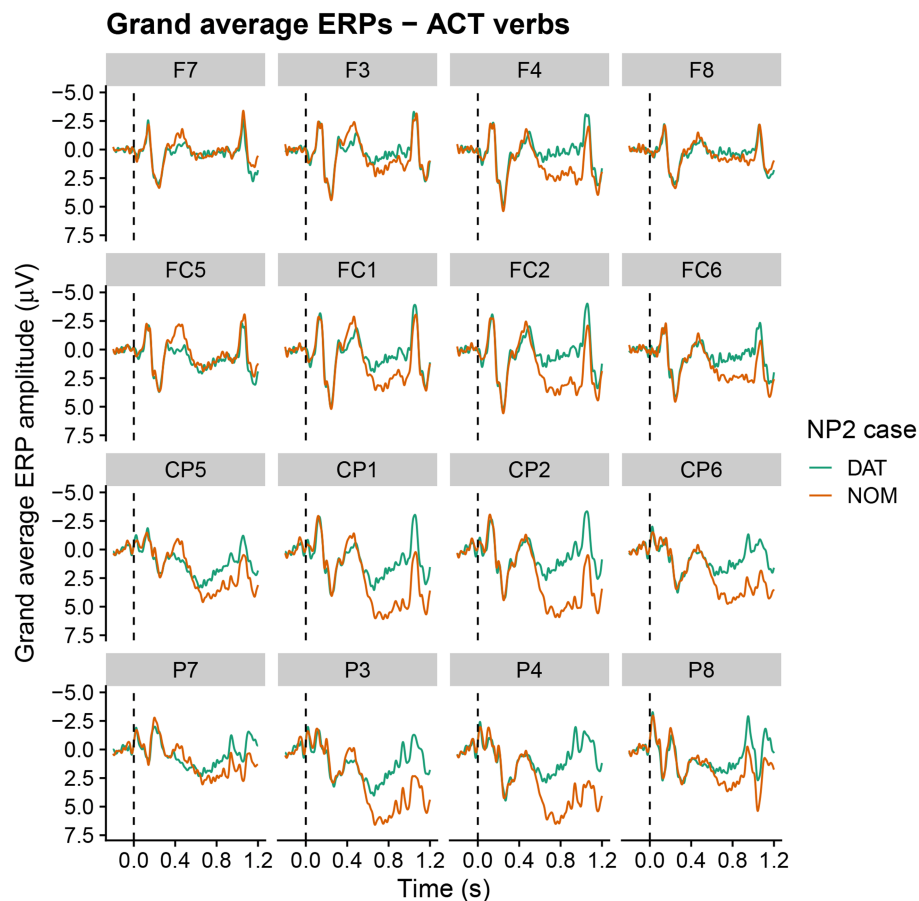


FIGURE 3 | Grand average ERPs at the critical NP2 position for sentences with active verbs (onset at the dashed vertical line). Negativity is plotted upwards.

(acceptable:1, unacceptable:0) as an additional fixed factor. In view of the restriction to only one type of verb, the factor verb was no longer included in the model (both fixed and random effects). All other parameters remained as described above for the general ERP models and models of this type were fit for both the N400 and late positivity time windows.

In line with our hypotheses, we focus on interactions of acceptability and case and, for each statistical model, interpret the highest-order interaction involving both of these predictors.

In the N400 time window, Wald tests (cf. **Table 4**) showed an interaction of case \times acceptability \times sagittality \times epoch, which is visualized and resolved in **Figure 8**. As is apparent from the figure, in spite of the interaction, there is no evidence for acceptability-based differences for either word order and this holds across the course of the experiment and for the different levels of sagittality.

For the late positivity time window, we observed an interaction of case \times acceptability \times epoch \times prestimulus amplitude (cf. **Table 5**). This interaction is visualized and resolved in **Figure 9**. Again, there is no evidence for acceptability-based differences and this pattern is broadly consistent across a range of prestimulus amplitudes (cf. **Figures S13–S15**).

In summary, there is no evidence that the ERP effects for the experiencer verbs vary on the basis of trial-by-trial changes in acceptability (full model summaries for the N400 and late positivity time windows are presented in **Tables S7–S10**).

4. DISCUSSION

We have presented an ERP experiment on Icelandic, with which we aimed to examine whether transitional processes of language change may be observable in the neural correlates of language comprehension prior to the change manifesting itself in overt, language-related behavior. The rationale behind this research question was that processes of language change affecting word order tend to arise from the need to process information that is increasingly ambiguously marked. In other words, if case marking is perceived as increasingly ambiguous, this can lead to a reinterpretation that in turn results in a stricter constituent order. We hypothesized that this type of reinterpretation should manifest itself in ERP responses during online language comprehension. If present, it would also constitute a highly interesting phenomenon at the interface between brain and behavior—both at the level of individual

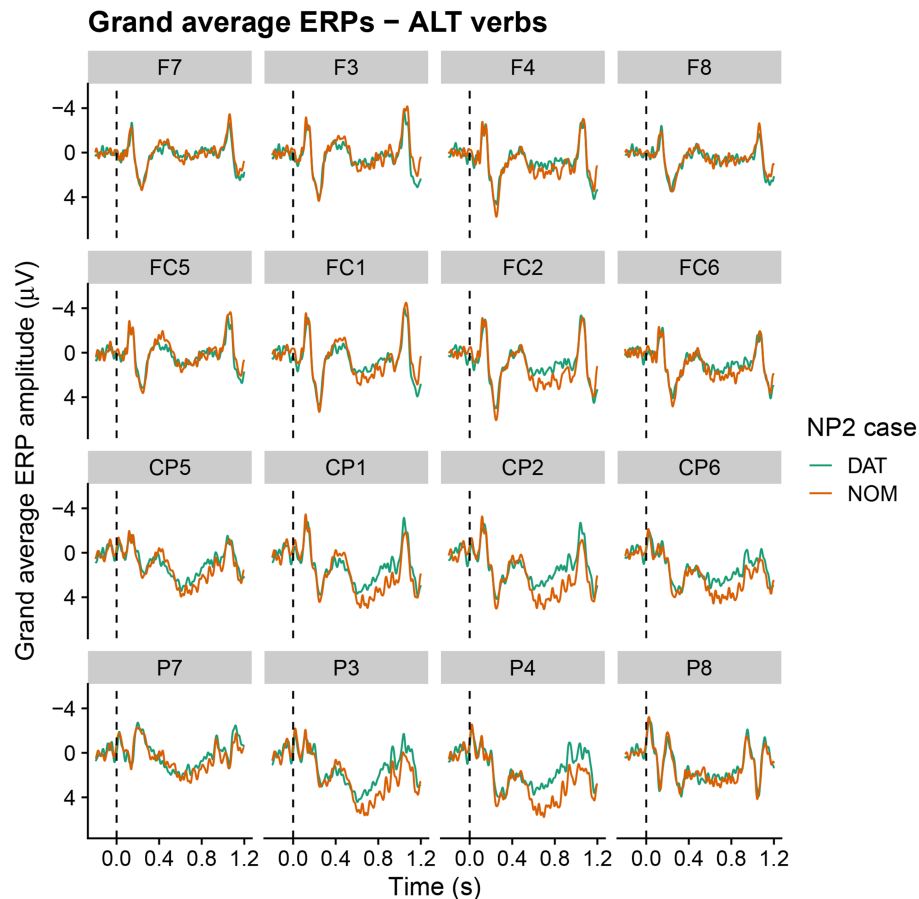


FIGURE 4 | Grand average ERPs at the critical NP2 position for sentences with alternating verbs (onset at the dashed vertical line). Negativity is plotted upwards.

speakers and in regard to the relation between neural processes, individual speaker behavior, and changes within communities of speakers.

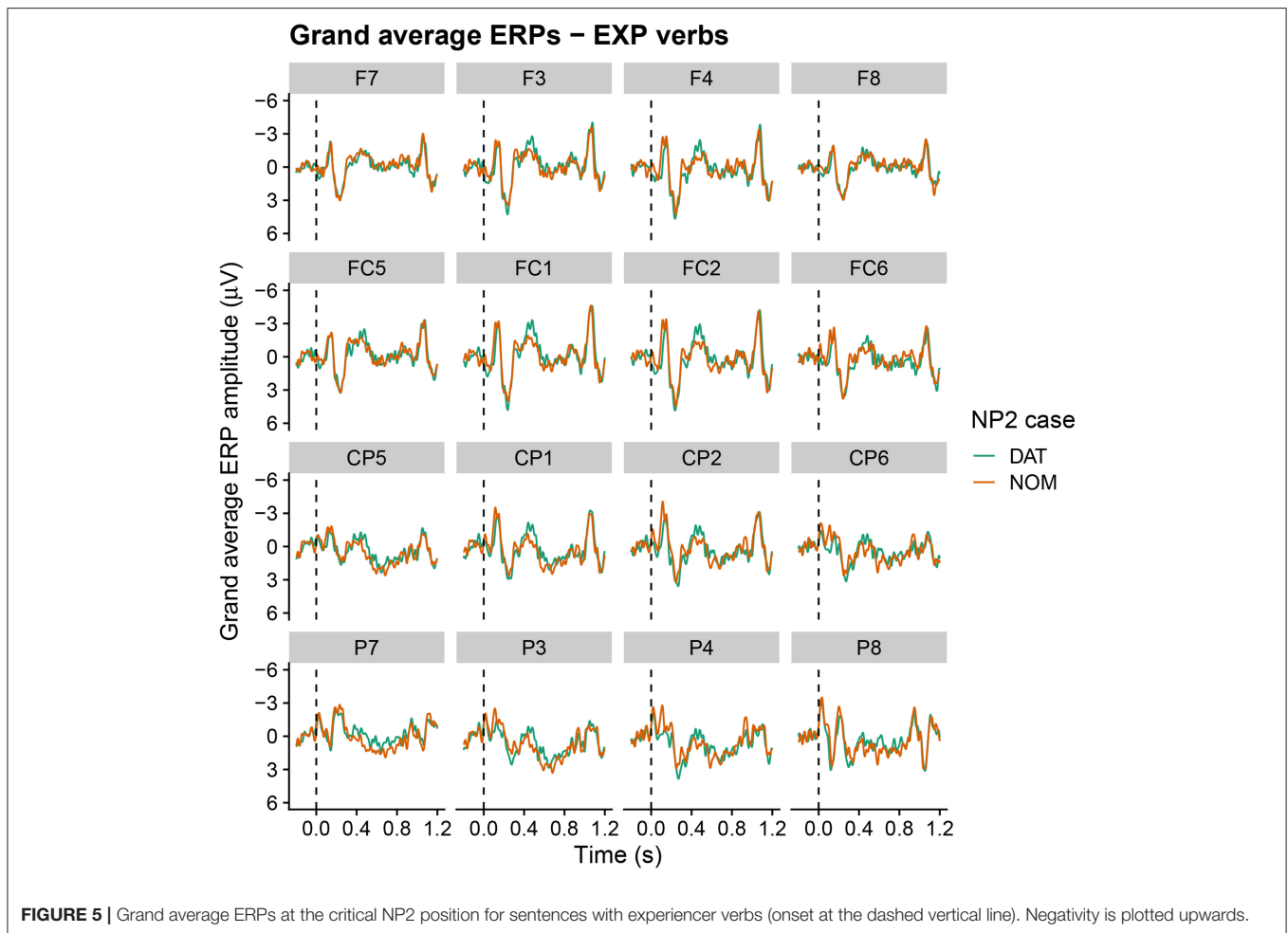
We indeed observed a pattern of results that was highly compatible with our hypotheses, i.e., a pattern suggesting that the transition from one grammar to another manifests itself in processing patterns at the neural level even before becoming apparent in overt language behavior (in the case of our study: assessment of sentence acceptability). In the following, we first summarize our results and explain why we believe they support this position. We then go on to discuss how the two ERP components observed—the N400 and late positivity—map onto behavior, before considering the implications of our findings for theories of language processing and language change.

4.1. Summary: Language Processing Precedes Language Change

As noted above, we contend that our results are consistent with the hypothesis that changes in language processing can precede overt language change. We base this claim on the 2-fold pattern of acceptability ratings and ERP patterns observed

in the present study. As we discuss in detail below, for each of the two verb types—alternating (ALT) and dative subject experiencer (EXP)—that we assume are undergoing a transition to the new target pattern (nominative subject, dative object), we observed a behavioral acceptability pattern that was “one step ahead” of what would be expected by the prescriptive grammar and an ERP pattern that was, in turn, one step ahead of the acceptability pattern.

Let us first consider the EXP verbs. Recall that, for these verbs, the prescriptive grammar requires dative subject and nominative object marking. From this perspective, they should thus be expected to show a pattern that is the mirror image of the one observed for active (ACT) verbs. However, while the behavioral ratings indeed show a higher acceptability for the dative-nominative (i.e., NP2 = nominative) as opposed to the nominative-dative (i.e., NP2 = dative) pattern for this verb class, the difference between the two patterns is not nearly as pronounced as the difference for nominative-dative vs. dative-nominative for ACT verbs (cf. **Figure 2**). In addition, EXP verbs also show highly variable judgement patterns across both participants and items (i.e., individual verbs; cf. **Figure 1**). This suggests that language change is already underway for



this verb class, with both individual speakers and individual verbs differing with regard to how far the change has already advanced⁴. Crucially, the ERP patterns observed for the EXP verbs are indicative of an even further advanced degree of change in that the prescriptively ungrammatical order conforming to the target state of Grammar B (nominative-dative) did not differ neurophysiologically from the grammatical (Grammar A) order dative-nominative. As both of these structures constitute an optimal realization in one of the two grammars, neither shows increased real-time processing costs relative to the other. This speaks in favor of a growing influence of Grammar B on the language comprehension architecture, in which it apparently already coexists with Grammar A for these particular structures—at least during online processing. We interpret the absence of differential ERP effects for this verb class as indicating that case marking has become relatively

⁴Note, however, that since we did not collect any production data from our participants, we do not know whether an acceptance of the nominative-dative pattern for EXP verbs also correlates with their use of this pattern in language production. It is therefore possible that, while at least some participants find this pattern highly acceptable, they do not yet produce it themselves.

uninformative for online interpretation. Hence, case marking patterns that are unexpected from the perspective of the current (prescriptive) Icelandic grammar—and even from the perspective of participants' own acceptability judgements—do not engender the typical ERP effects that are known to accompany these mismatches (N400, late positivity). The response-contingent analysis of the trial-by-trial ERP responses to experiencer verbs further supports this interpretation by demonstrating that the apparent absence of an effect cannot be explained by a trial-by-trial fluctuation of ERP responses depending on whether the construction was judged to be acceptable on a particular trial or not (i.e., it was not the case that sentences judged to be unacceptable engendered an N400-late positivity response irrespective of the case marking pattern). Our interpretation that case marking is no longer informative for online argument interpretation in these types of experiencer constructions in Icelandic is additionally corroborated by the observed pattern of comprehension accuracy, which was generally lower than that for the other two verb classes and did not differ depending on word order.

For the ALT verbs, the transition toward Grammar B is already much further advanced. Despite the possible grammatical

TABLE 2 | Summary of effects in N400 time window (Type II Wald Tests).

	Chisq	Df	Pr(> Chisq)
scale(prestim)	2643.183	1	0.000
verb	10.934	2	0.004
case	14.302	1	0.000
lat.	245.452	1	0.000
sag.	694.042	1	0.000
scale(epoch)	7.140	1	0.008
scale(prestim):verb	93.725	2	0.000
scale(prestim):case	8.088	1	0.004
verb:case	9.224	2	0.010
scale(prestim):lat.	23.091	1	0.000
verb:lat.	1.233	2	0.540
case:lat.	5.586	1	0.018
scale(prestim):sag.	107.633	1	0.000
verb:sag.	46.735	2	0.000
case:sag.	62.627	1	0.000
lat.:sag.	2.973	1	0.085
scale(prestim):scale(epoch)	22.067	1	0.000
verb:scale(epoch)	17.447	2	0.000
case:scale(epoch)	5.012	1	0.025
lat.:scale(epoch)	2.632	1	0.105
sag.:scale(epoch)	23.460	1	0.000
scale(prestim):verb:case	4.781	2	0.092
scale(prestim):verb:lat.	0.549	2	0.760
scale(prestim):case:lat.	6.555	1	0.010
verb:case:lat.	1.459	2	0.482
scale(prestim):verb:sag.	2.466	2	0.291
scale(prestim):case:sag.	15.001	1	0.000
verb:case:sag.	37.054	2	0.000
scale(prestim):lat.:sag.	0.576	1	0.448
verb:lat.:sag.	0.018	2	0.991
case:lat.:sag.	0.008	1	0.931
scale(prestim):verb:scale(epoch)	17.295	2	0.000
scale(prestim):case:scale(epoch)	2.388	1	0.122
verb:case:scale(epoch)	6.376	2	0.041
scale(prestim):lat.:scale(epoch)	0.164	1	0.685
verb:lat.:scale(epoch)	0.039	2	0.981
case:lat.:scale(epoch)	0.761	1	0.383
scale(prestim):sag.:scale(epoch)	11.177	1	0.001
verb:sag.:scale(epoch)	1.297	2	0.523
case:sag.:scale(epoch)	0.687	1	0.407
lat.:sag.:scale(epoch)	0.112	1	0.737
scale(prestim):verb:case:lat.	0.246	2	0.884
scale(prestim):verb:case:sag.	1.474	2	0.478
scale(prestim):verb:lat.:sag.	1.391	2	0.499
scale(prestim):case:lat.:sag.	0.702	1	0.402
verb:case:lat.:sag.	0.506	2	0.776
scale(prestim):verb:case:scale(epoch)	14.839	2	0.001
scale(prestim):verb:lat.:scale(epoch)	2.729	2	0.255

(Continued)

TABLE 2 | Continued

	Chisq	Df	Pr(> Chisq)
scale(prestim):case:lat.:scale(epoch)	4.570	1	0.033
verb:case:lat.:scale(epoch)	2.832	2	0.243
scale(prestim):verb:sag.:scale(epoch)	2.489	2	0.288
scale(prestim):case:sag.:scale(epoch)	0.029	1	0.865
verb:case:sag.:scale(epoch)	15.361	2	0.000
scale(prestim):lat.:sag.:scale(epoch)	0.010	1	0.919
verb:lat.:sag.:scale(epoch)	0.401	2	0.818
case:lat.:sag.:scale(epoch)	0.031	1	0.861
scale(prestim):verb:case:lat.:sag.	0.348	2	0.840
scale(prestim):verb:case:lat.:scale(epoch)	0.895	2	0.639
scale(prestim):verb:case:sag.:scale(epoch)	9.679	2	0.008
scale(prestim):verb:lat.:sag.:scale(epoch)	0.079	2	0.961
scale(prestim):case:lat.:sag.:scale(epoch)	2.172	1	0.141
verb:case:lat.:sag.:scale(epoch)	0.069	2	0.966
scale(prestim):verb:case:lat.:sag.:scale(epoch)	2.233	2	0.327

dative-before-nominative realization (licensed by Grammar A), these verbs show a very similar and only slightly weaker neurophysiological response to that for the ACT verbs, in which the dative-before-nominative order is completely ruled out. Even though Grammar B obviously already dominates the processing of these structures, the weaker disadvantage for the dative-initial word order in comparison to the active verbs reflects the remaining remnants of Grammar A's influence, as does the higher degree of by-participant and by-item variability for ALT verbs (cf. **Figures S3, S4**). Strikingly, while the alternating verbs show no difference between word orders in the N400 at the beginning of the experiment, they converge on the pattern shown by the active verbs (increased N400 amplitude for dative-nominative vs. nominative-dative orders) by the end of the experimental session (see **Figure 4**). We take this to reflect the higher degree of uncertainty surrounding the dominant or preferred structure with these verbs in comparison to active verbs. Supporting this notion, there is a high degree of judgement variability for the dative-nominative pattern with alternating verbs (**Figure 1**)—paralleling that for the experiencer verbs. The nominative-dative order, by contrast, is consistently judged as acceptable, thus patterning with the results for the active verbs. Comprehension accuracy mirrors these results in that participants were highly accurate in responding to the comprehension questions for alternating verbs with nominative-dative orders, but considerably less accurate in the case of dative-nominative orders.

Finally, the active verbs showed a highly consistent pattern across all the measures employed here, as was expected given that they already conform to the requirements of the target grammar (B). These verbs showed a clear N400—late positivity pattern for dative-nominative vs. nominative-dative orders, which was apparent across the entire experiment. Nominative-dative orders were consistently judged to be acceptable across participants and items, while dative-nominative

TABLE 3 | Summary of effects in Late Positivity time window (Type II Wald Tests).

	Chisq	Df	Pr(> Chisq)
scale(prestim)	30281.850	1	0.000
verb	10.227	2	0.006
case	10.718	1	0.001
lat.	55.111	1	0.000
sag.	147.238	1	0.000
scale(epoch)	12.656	1	0.000
scale(prestim):verb	126.364	2	0.000
scale(prestim):case	0.189	1	0.664
verb:case	7.516	2	0.023
scale(prestim):lat.	26.950	1	0.000
verb:lat.	2.196	2	0.334
case:lat.	0.487	1	0.485
scale(prestim):sag.	112.018	1	0.000
verb:sag.	48.394	2	0.000
case:sag.	113.850	1	0.000
lat.:sag.	37.826	1	0.000
scale(prestim):scale(epoch)	11.877	1	0.001
verb:scale(epoch)	7.696	2	0.021
case:scale(epoch)	5.780	1	0.016
lat.:scale(epoch)	0.403	1	0.526
sag.:scale(epoch)	3.341	1	0.068
scale(prestim):verb:case	27.713	2	0.000
scale(prestim):verb:lat.	1.539	2	0.463
scale(prestim):case:lat.	0.414	1	0.520
verb:case:lat.	0.966	2	0.617
scale(prestim):verb:sag.	5.904	2	0.052
scale(prestim):case:sag.	1.541	1	0.215
verb:case:sag.	20.950	2	0.000
scale(prestim):lat.:sag.	0.138	1	0.710
verb:lat.:sag.	0.141	2	0.932
case:lat.:sag.	0.510	1	0.475
scale(prestim):verb:scale(epoch)	97.860	2	0.000
scale(prestim):case:scale(epoch)	0.120	1	0.729
verb:case:scale(epoch)	1.782	2	0.410
scale(prestim):lat.:scale(epoch)	0.002	1	0.964
verb:lat.:scale(epoch)	1.694	2	0.429
case:lat.:scale(epoch)	0.523	1	0.469
scale(prestim):sag.:scale(epoch)	4.024	1	0.045
verb:sag.:scale(epoch)	6.987	2	0.030
case:sag.:scale(epoch)	2.983	1	0.084
lat.:sag.:scale(epoch)	0.028	1	0.867
scale(prestim):verb:case:lat.	0.043	2	0.979
scale(prestim):verb:case:sag.	0.599	2	0.741
scale(prestim):verb:lat.:sag.	0.281	2	0.869
scale(prestim):case:lat.:sag.	0.020	1	0.887
verb:case:lat.:sag.	0.169	2	0.919
scale(prestim):verb:case:scale(epoch)	16.156	2	0.000
scale(prestim):verb:lat.:scale(epoch)	2.839	2	0.242

(Continued)

TABLE 3 | Continued

	Chisq	Df	Pr(> Chisq)
scale(prestim):case:lat.:scale(epoch)	4.166	1	0.041
verb:case:lat.:scale(epoch)	4.350	2	0.114
scale(prestim):verb:sag.:scale(epoch)	6.401	2	0.041
scale(prestim):case:sag.:scale(epoch)	0.164	1	0.686
verb:case:sag.:scale(epoch)	0.837	2	0.658
scale(prestim):lat.:sag.:scale(epoch)	0.651	1	0.420
verb:lat.:sag.:scale(epoch)	0.658	2	0.720
case:lat.:sag.:scale(epoch)	0.618	1	0.432
scale(prestim):verb:case:lat.:sag.	0.146	2	0.930
scale(prestim):verb:case:lat.:scale(epoch)	1.014	2	0.602
scale(prestim):verb:case:sag.:scale(epoch)	1.385	2	0.500
scale(prestim):verb:lat.:sag.:scale(epoch)	0.040	2	0.980
scale(prestim):case:lat.:sag.:scale(epoch)	2.724	1	0.099
verb:case:lat.:sag.:scale(epoch)	0.546	2	0.761
scale(prestim):verb:case:lat.:sag.:scale(epoch)	2.591	2	0.274

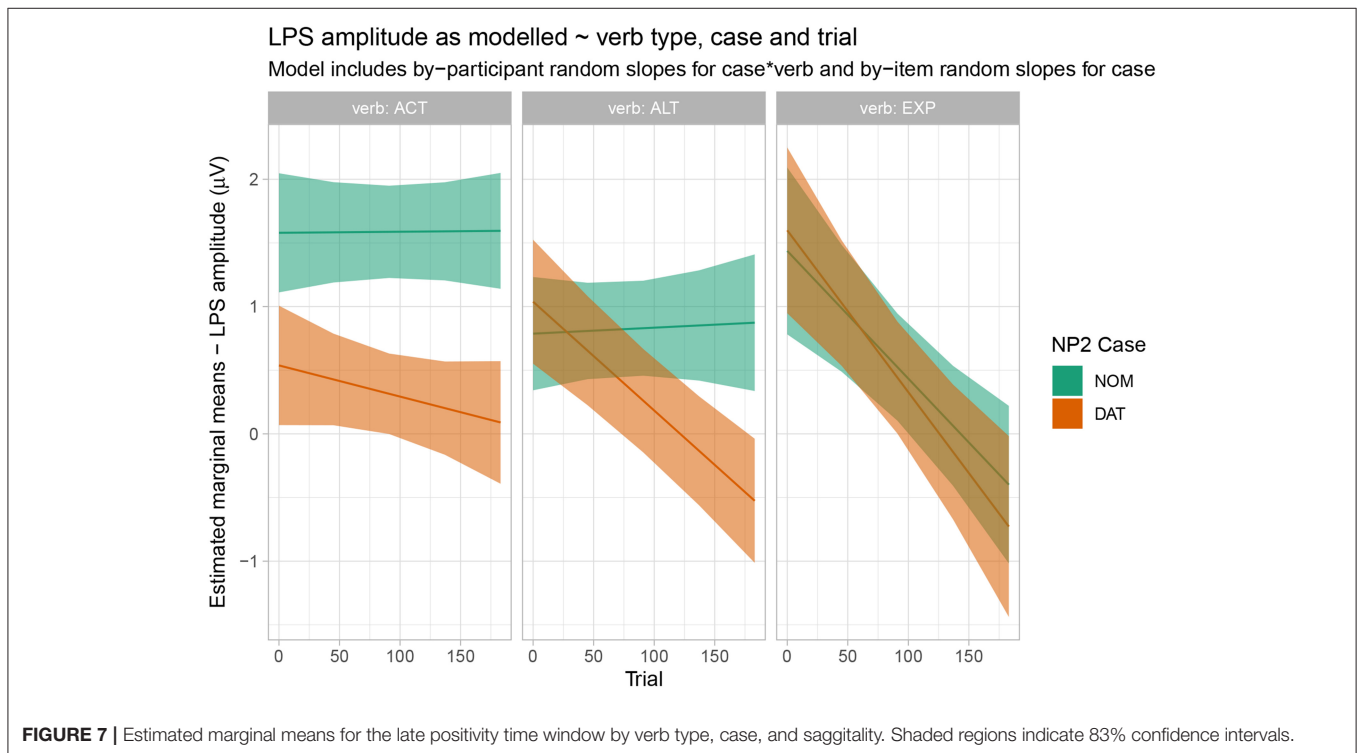
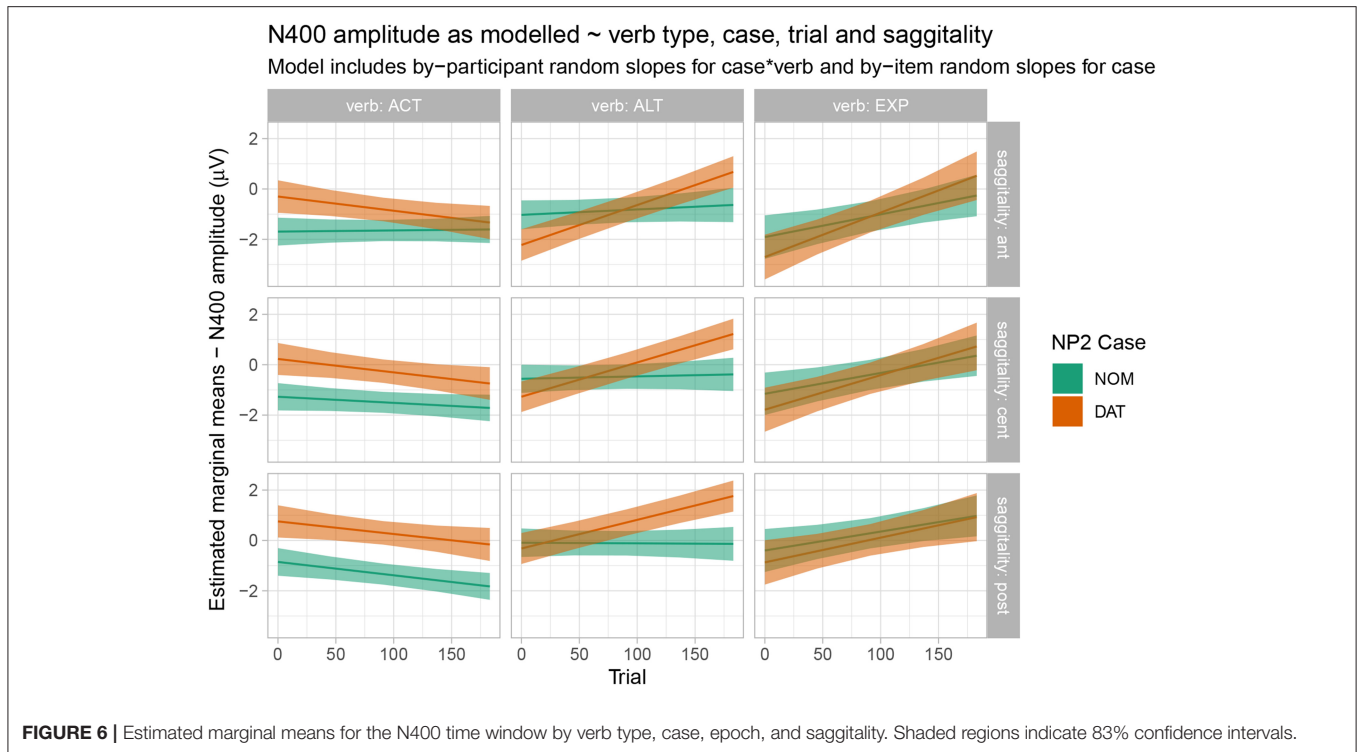
orders were consistently rejected. Intriguingly, the results of the comprehension task revealed that sentences with active verbs were comprehended highly accurately independently of the word order. This was the case in spite of the low acceptability of the dative-nominative order. We interpret this pattern as being indicative of low interpretative value of case marking in these structures: all that matters for comprehension is which argument occupies the subject position. This is reminiscent of how language comprehension operates in modern English, in which word order always dominates morphological marking as an interpretative cue.

4.2. The Relation Between the N400 and Late Positivity Components and Behavior

Having discussed our interpretation of the overall pattern of results, we now turn to a more mechanistic account of what we consider the N400 and late positivity components to reflect in the current data.

4.2.1. N400

We have recently proposed that N400 effects reflect precision-weighted prediction errors (Bornkessel-Schlesewsky and Schlewsky, 2019) in the sense of a predictive-coding account of brain function (cf. Friston, 2005, 2010). In brief, predictive coding assumes that the brain actively constructs explanations for its sensory input and that this involves maintaining an internal generative (predictive) model of the world around us. The brain is thus constantly engaged in generating predictions for upcoming sensory input and in matching these to the input actually encountered. Prediction errors (i.e., mismatches between prediction and input) can lead to internal model updating. Crucially, predictions differ in regard to their precision, which is defined as the inverse of variance and thus essentially reflects the degree of (un)certainly (Feldman and Friston, 2010). Prediction precision has been shown to modulate mismatch negativity (MMN) effects (Todd et al., 2014) and,



as posited in Bornkessel-Schlesewsky and Schlesewsky (2019), there is evidence to suggest that the same holds for N400 effects in language. From this perspective, we would expect to observe

more pronounced N400 effects for higher precision predictions. This approach constitutes a promising conceptual framework for interpreting the N400 effects in the current experiment (for

TABLE 4 | Summary of experiencer verb analysis including acceptability effects in N400 time window (Type II Wald Tests).

	Chisq	Df	Pr(> Chisq)
scale(prestim)	206.045	1	0.000
case	0.106	1	0.745
accept	2.781	1	0.095
lat.	69.315	1	0.000
sag.	293.656	1	0.000
scale(epoch)	9.880	1	0.002
scale(prestim):case	0.070	1	0.791
scale(prestim):accept	15.022	1	0.000
case:accept	0.701	1	0.402
scale(prestim):lat.	2.780	1	0.095
case:lat.	0.038	1	0.844
accept:lat.	0.276	1	0.599
scale(prestim):sag.	22.032	1	0.000
case:sag.	2.325	1	0.127
accept:sag.	0.851	1	0.356
lat.:sag.	0.701	1	0.402
scale(prestim):scale(epoch)	2.949	1	0.086
case:scale(epoch)	0.287	1	0.592
accept:scale(epoch)	4.888	1	0.027
lat.:scale(epoch)	0.146	1	0.703
sag.:scale(epoch)	15.625	1	0.000
scale(prestim):case:accept	34.062	1	0.000
scale(prestim):case:lat.	1.224	1	0.269
scale(prestim):accept:lat.	0.088	1	0.767
case:accept:lat.	1.484	1	0.223
scale(prestim):case:sag.	3.926	1	0.048
scale(prestim):accept:sag.	5.849	1	0.016
case:accept:sag.	6.477	1	0.011
scale(prestim):lat.:sag.	0.177	1	0.674
case:lat.:sag.	0.491	1	0.483
accept:lat.:sag.	0.954	1	0.329
scale(prestim):case:scale(epoch)	0.805	1	0.370
scale(prestim):accept:scale(epoch)	58.182	1	0.000
case:accept:scale(epoch)	8.719	1	0.003
scale(prestim):lat.:scale(epoch)	0.604	1	0.437
case:lat.:scale(epoch)	3.491	1	0.062
accept:lat.:scale(epoch)	0.143	1	0.705
scale(prestim):sag.:scale(epoch)	1.119	1	0.290
case:sag.:scale(epoch)	3.957	1	0.047
accept:sag.:scale(epoch)	0.427	1	0.514
lat.:sag.:scale(epoch)	0.002	1	0.963
scale(prestim):case:accept:lat.	0.753	1	0.386
scale(prestim):case:accept:sag.	0.239	1	0.625
scale(prestim):case:lat.:sag.	0.034	1	0.853
scale(prestim):accept:lat.:sag.	0.917	1	0.338
case:accept:lat.:sag.	0.635	1	0.426
scale(prestim):case:accept:scale(epoch)	5.886	1	0.015

(Continued)

TABLE 4 | Continued

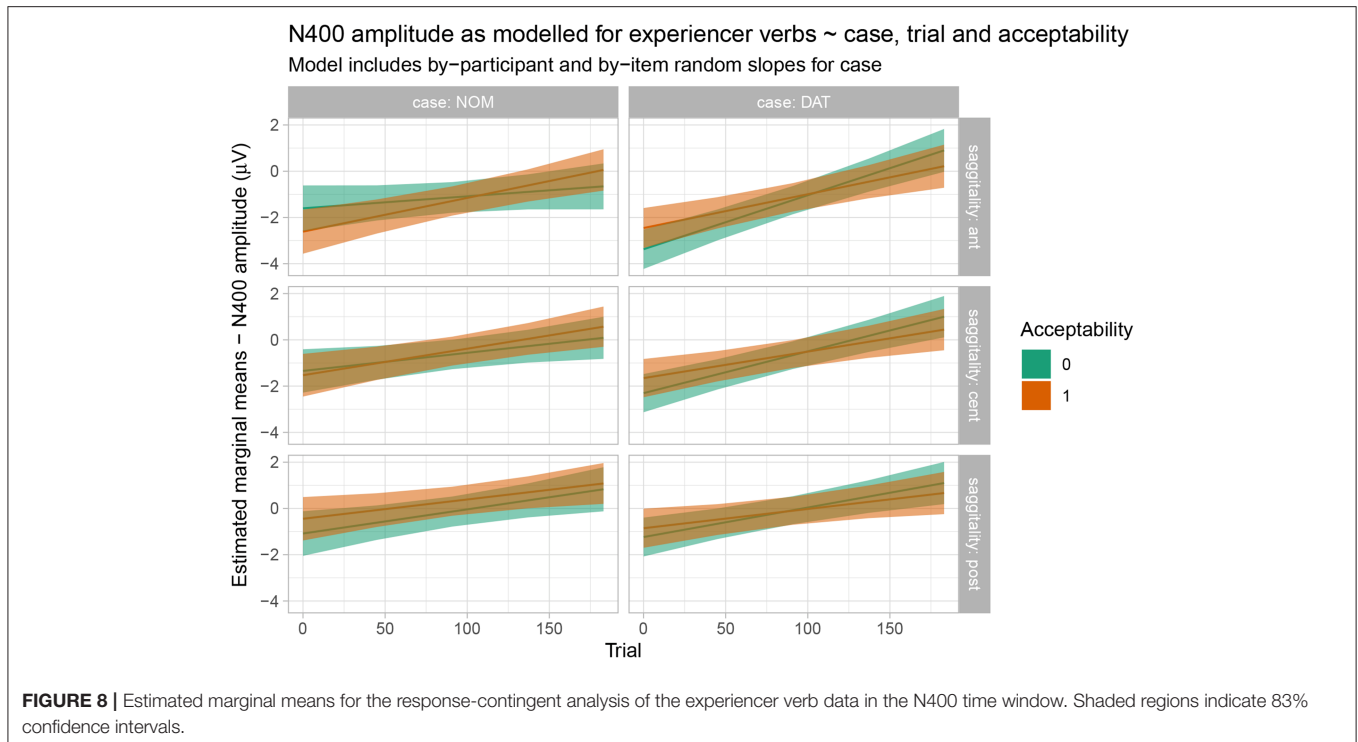
	Chisq	Df	Pr(> Chisq)
scale(prestim):case:lat.:scale(epoch)	0.044	1	0.833
scale(prestim):accept:lat.:scale(epoch)	3.992	1	0.046
case:accept:lat.:scale(epoch)	0.013	1	0.909
scale(prestim):case:sag.:scale(epoch)	0.434	1	0.510
scale(prestim):accept:sag.:scale(epoch)	4.707	1	0.030
case:accept:sag.:scale(epoch)	5.816	1	0.016
scale(prestim):lat.:sag.:scale(epoch)	0.095	1	0.758
case:lat.:sag.:scale(epoch)	0.014	1	0.906
accept:lat.:sag.:scale(epoch)	0.200	1	0.655
scale(prestim):case:accept:lat.:sag.	0.001	1	0.977
scale(prestim):case:accept:lat.:scale(epoch)	1.219	1	0.270
scale(prestim):case:accept:sag.:scale(epoch)	0.256	1	0.613
scale(prestim):case:lat.:sag.:scale(epoch)	0.058	1	0.810
scale(prestim):accept:lat.:sag.:scale(epoch)	0.156	1	0.693
case:accept:lat.:sag.:scale(epoch)	0.003	1	0.957
scale(prestim):case:accept:lat.:sag.:scale(epoch)	0.000	1	0.997

a comparison to other current interpretations of the N400, see Bornkessel-Schlesewsky and Schlesewsky, 2019).

In sentences with active verbs, the language comprehension system is able to generate a high-precision prediction for a post-verbal dative argument. When this prediction is not borne out, the resulting prediction error is reflected in an N400 effect. The prediction (and precision of the prediction) is highly stable, thus leading to comparable N400 effects across the course of the experiment for active verbs.

For alternating verbs, the situation is more complex. While the nominative-dative order is highly acceptable across the board, it has a competitor in the dative-nominative order—with the degree of competition varying across participants and items. Accordingly, there is a lower precision prediction for the case marking of the post-verbal NP and no N400 difference at the beginning of experiment. Across the course of the experiment, however, the precision of the prediction for nominative-dative appears to strengthen, and an N400 effect emerges. We speculate that this by-trial change may have been precipitated by the presence of a high number of active verbs in the experiment. (But note that there was no comparable emergence of an N400 effect for the experiencer verbs, thus suggesting that alternating verbs were more strongly susceptible to such an influence). Yet whatever the explanation for the emergence of an N400 effect for dative-nominative vs. nominative-dative orders, this pattern attests to a less stable pattern than that for the active verbs, as also seen in the behavioral data. For the alternating verbs, uncertainty arising from the variability in the dative-nominative order is key to the overall pattern of results.

The experiencer verbs show a high behavioral uncertainty for both word orders. Thus, predictions in online processing are of a very low precision and this



manifests itself in the absence of reliable N400 effects in either direction⁵.

4.2.2. Late Positivity

The late positivity effects in the current experiment showed a similar pattern to those observed in the N400: active verbs showed a positivity for dative-nominative vs. nominative-dative orders across the entire experiment; for alternating verbs, a similar effect emerged over the course of the experiment; experiencer verbs showed no differential late positivity effects. Overall, the late positivity appears to reflect the dominant acceptability pattern for each verb class: a clear preference for nominative-dative for active verbs; a similar, but weaker preference for alternating verbs; and high variability for experiencer verbs. In spite of the generally similar patterns for the N400 and late positivity, we expected that the late positivity effects observed should be tied more strongly to

the overall evaluation of the structures in question than to their incremental comprehension (and the prediction-based effects involved therein). We derive this assumption from the proposal that late positivity effects in language should be viewed as members of the P300 family (e.g., Coulson et al., 1998; Sassenhagen et al., 2014) and that they are therefore connected more closely to the motivational salience of a stimulus and how this translates to behavior (for discussion in comparison to the N400, see Bornkessel-Schlesewsky and Schlesewsky, 2019).

In order to test this assumption further, we computed two generalized linear mixed models, in which we examined the extent to which single-trial N400 and late positivity amplitudes can predict single-trial acceptability ratings. We included (z-transformed) mean amplitude for the respective time window, laterality and sagittality in the model as fixed effects, with random intercepts grouped by participant and item. Both the N400 and the late positivity model fits were improved by additionally adding verb type as a predictor [likelihood ratio test for N400 model: $\chi^2(16) = 86.88$, $p < 0.001$; late positivity model: $\chi^2(16) = 262.55$, $p < 0.001$]. While both N400 and late positivity amplitudes predicted acceptability on a single trial basis [N400 amplitude x verb type: $\chi^2(2) = 59.76$, $p < 0.001$; LPS amplitude x verb type: $\chi^2(2) = 218.02$, $p < 0.001$], the late positivity model showed an overall better fit to the data (AIC for N400 model including verb type: 114407; AIC for late positivity model including verb type: 113320). In addition, as shown in **Figures S16, S17**, late positivity amplitudes showed a stronger relationship with acceptability than N400 amplitudes.

⁵We suggest that the different patterns for the experiencer vs. alternating verbs may reflect the fact that only the experiencer verbs are subject to two competing patterns: nominative-dative, which reflects the canonical subject-object pattern in terms of grammatical relations, and dative-nominative, which reflects the semantic role hierarchy Experiencer > Stimulus. In other words, the association between dative case marking and the Experiencer role serves to bolster the dative-nominative pattern for the experiencer verbs—this is likely also the mechanism underlying dative substitution for experiencer verbs with accusative subjects, cf. example 1. Thus, while the experiencer verbs are subject to a high level of uncertainty due to two competing case marking patterns based on grammatical relations and semantic roles, respectively, the dative-nominative pattern for alternating verbs is not supported by semantic information and thus more susceptible to change over the course of the experiment.

TABLE 5 | Summary of experiencer verb analysis including acceptability effects in late positivity time window (Type II Wald Tests).

	Chisq	Df	Pr(> Chisq)
scale(prestim)	8192.291	1	0.000
case	0.350	1	0.554
accept	13.170	1	0.000
lat.	16.012	1	0.000
sag.	8.666	1	0.003
scale(epoch)	15.275	1	0.000
scale(prestim):case	0.459	1	0.498
scale(prestim):accept	6.446	1	0.011
case:accept	8.356	1	0.004
scale(prestim):lat.	1.835	1	0.176
case:lat.	0.111	1	0.739
accept:lat.	0.075	1	0.784
scale(prestim):sag.	17.865	1	0.000
case:sag.	5.357	1	0.021
accept:sag.	0.615	1	0.433
lat.:sag.	10.965	1	0.001
scale(prestim):scale(epoch)	57.758	1	0.000
case:scale(epoch)	1.331	1	0.249
accept:scale(epoch)	24.118	1	0.000
lat.:scale(epoch)	0.266	1	0.606
sag.:scale(epoch)	2.820	1	0.093
scale(prestim):case:accept	91.025	1	0.000
scale(prestim):case:lat.	0.363	1	0.547
scale(prestim):accept:lat.	0.263	1	0.608
case:accept:lat.	0.275	1	0.600
scale(prestim):case:sag.	2.001	1	0.157
scale(prestim):accept:sag.	1.847	1	0.174
case:accept:sag.	0.009	1	0.925
scale(prestim):lat.:sag.	0.018	1	0.894
case:lat.:sag.	0.048	1	0.826
accept:lat.:sag.	1.170	1	0.279
scale(prestim):case:scale(epoch)	0.794	1	0.373
scale(prestim):accept:scale(epoch)	37.292	1	0.000
case:accept:scale(epoch)	0.157	1	0.692
scale(prestim):lat.:scale(epoch)	0.105	1	0.746
case:lat.:scale(epoch)	5.341	1	0.021
accept:lat.:scale(epoch)	2.350	1	0.125
scale(prestim):sag.:scale(epoch)	4.474	1	0.034
case:sag.:scale(epoch)	0.175	1	0.676
accept:sag.:scale(epoch)	2.593	1	0.107
lat.:sag.:scale(epoch)	0.146	1	0.702
scale(prestim):case:accept:lat.	1.059	1	0.303
scale(prestim):case:accept:sag.	2.148	1	0.143
scale(prestim):case:lat.:sag.	0.054	1	0.816
scale(prestim):accept:lat.:sag.	0.649	1	0.421
case:accept:lat.:sag.	0.114	1	0.735
scale(prestim):case:accept:scale(epoch)	4.018	1	0.045

(Continued)

TABLE 5 | Continued

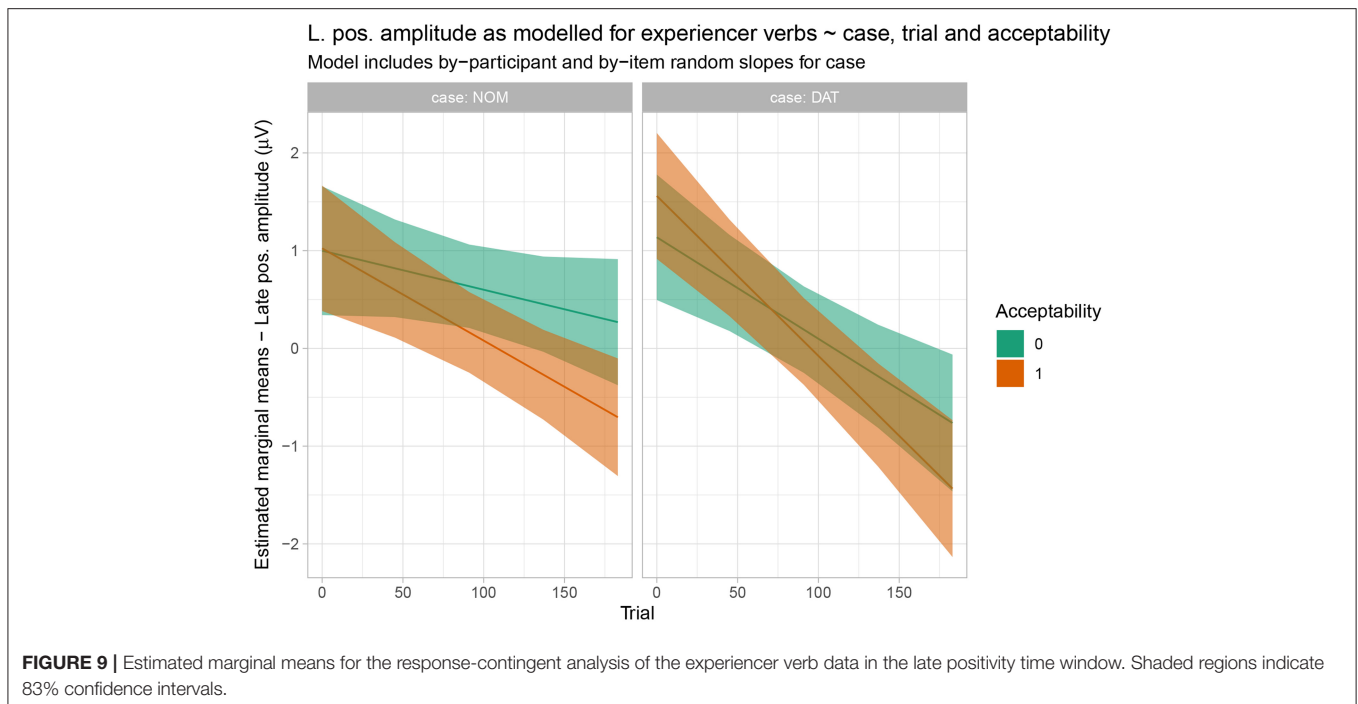
	Chisq	Df	Pr(> Chisq)
scale(prestim):case:lat.:scale(epoch)	0.481	1	0.488
scale(prestim):accept:lat.:scale(epoch)	3.599	1	0.058
case:accept:lat.:scale(epoch)	2.389	1	0.122
scale(prestim):case:sag.:scale(epoch)	0.004	1	0.951
scale(prestim):accept:sag.:scale(epoch)	2.486	1	0.115
case:accept:sag.:scale(epoch)	1.164	1	0.281
scale(prestim):lat.:sag.:scale(epoch)	0.336	1	0.562
case:lat.:sag.:scale(epoch)	0.021	1	0.884
accept:lat.:sag.:scale(epoch)	0.288	1	0.591
scale(prestim):case:accept:lat.:sag.	0.000	1	0.983
scale(prestim):case:accept:lat.:scale(epoch)	0.002	1	0.966
scale(prestim):case:accept:sag.:scale(epoch)	0.018	1	0.892
scale(prestim):case:lat.:sag.:scale(epoch)	0.083	1	0.773
scale(prestim):accept:lat.:sag.:scale(epoch)	1.786	1	0.181
case:accept:lat.:sag.:scale(epoch)	0.649	1	0.420
scale(prestim):case:accept:lat.:sag.:scale(epoch)	0.919	1	0.338

Interestingly, in both times windows, EEG amplitudes were more strongly predictive of acceptability for active and alternating than for experiencer verbs, thus further supporting our argument of highly variable EEG responses for experiencer verbs that are not correlated with acceptability.

In summary, single-trial late positivity amplitudes were more predictive of behavior (acceptability) than N400 amplitudes, as expected. Thus, in spite of the fact that the late positivity effects observed here showed larger amplitudes than the N400 effects, we suggest that the N400 effects will be more predictive of language change due to their higher sensitivity to the demands of online comprehension and stronger independence from behavior. Whether this assumption is indeed correct, however, cannot be determined on the basis of the present findings, since our study does not include any longitudinal or diachronic data. It should therefore be viewed as a testable hypothesis for future research based on the current results, rather than as a conclusion from the current study.

4.3. Implications for The Relation Between Language Processing and Language Change

These findings provide initial converging evidence for an intriguing picture of the dynamics of language change. In particular, they suggest that we can identify three successively less conservative levels of language behavior: (a) the prescriptive grammar and conscious behavior adhering to its rules; (b) the intuitions of native speakers under time pressure—and thereby under similar circumstances as in real life communication; and (c) the underlying source of all of these behavioral responses: the human brain. These three dimensions are ordered hierarchically with respect to one another, such that each is “one step ahead”



of the previous stage⁶. While it is well-known that changes in prescriptive grammar result from an adaptation to transitions that have already been established in everyday language use, our findings suggest that the neural processing architecture in turn paves the way for these changes in overt language-based behavior. Brain responses—which, as discussed in the introduction, can be viewed as reflecting the reinterpretation processes that foreshadow at least certain processes of language change—can therefore be used as early indicators for transitions that will subsequently emerge in first the informal and later the formal (normative) uses of a particular language. Depending on the particular neurophysiological patterns observed, concrete predictions for the direction of language change can be formulated.

Regarding Icelandic, our data suggest that the alternating verbs will come to be associated with a fixed nominative-initial word order, thereby completing a change that is already relatively far advanced. More interestingly, the dative subject-experienter verbs are predicted to first turn into alternating verbs in both surface behavior and prescriptive grammar, before following the current alternating verbs on their path toward the fixed nominative-first active constructions. As a consequence, dative subjects in Icelandic will become first an endangered and

subsequently an extinct species. This will likely be the starting point for a complete deconstruction of the morphological system.

As noted above, we suggest that N400 effects may be particularly promising early indicators of the initial stages of such a process, namely reinterpretation during language processing. The proposal that N400 effects reflect precision-weighted prediction errors provides a neurobiological grounding for this claim: as an information source becomes more ambiguous, it becomes less reliable for formulating predictions and any predictions generated during online comprehension are thus of lower precision. Reduced N400 effects to structures that are incompatible with the current prescriptive grammar could thus provide us with an early “snapshot of the brain in transition” and hence the capacity to predict the directions that languages will take in their future development.

DATA AVAILABILITY STATEMENT

Datasets are in a publicly accessible repository: The datasets generated for this study and the analysis code can be found in an Open Science Framework repository, <https://osf.io/zp6yv/>.

ETHICS STATEMENT

The present study was performed in accordance with the ethical standards laid down in the Declaration of Helsinki. Participants gave written informed consent before the beginning of the experiment and were informed that they could discontinue the study at any time should they wish to do so. The experimental protocols were approved by the ethics committee of the Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany.

⁶Note that this assumption is further supported by the phenomenon of “case sickness” in Icelandic (cf. Eythórrsson, 2000). Case sickness refers, for example, to dative case marking being used to mark the subjects of accusative subject verbs, a tendency that is common in Icelandic but frowned upon by prescriptive grammars. The motivation for this change is likely that dative is commonly used to mark actor-type arguments, while accusative is not. Hence, language production is one step ahead of the prescriptive grammar. At the same time, our results suggest that the change is already further advanced in language comprehension in that position is weighted more strongly than case for interpretation, possibly due to the reduction of the number of cases permissible in particular structural positions.

AUTHOR'S NOTE

We are indebted to Jörgen Pind for the opportunity to acquire data at the Department of Psychology of the University of Iceland, Reykjavik. We would further like to thank Thórhallur Eythórsson for helpful discussions and Ella Björt Teague and Sigrúnsif Jóelsdóttir for assistance in data acquisition. This manuscript was written in RMarkdown using the R package *papaja* (Aust and Barth, 2018).

AUTHOR CONTRIBUTIONS

IB-S, MS, and DR designed the research. DR conducted the research. IB-S and DR analyzed the data. IB-S, MS, RM, and DR wrote the paper.

REFERENCES

- Alday, P. M. (2018). *lmerOut: LaTeX Output for Mixed Effects Models with lme4*. Available online at: <https://bitbucket.org/palday/lmerout/src/master/>
- Alday, P. M. (2019). How much baseline correction do we need in ERP research? Extended GLM model can replace baseline correction while lifting its limits. *Psychophysiology* 56, 1–14. doi: 10.1111/psyp.13451
- Alday, P. M., and Kretzschmar, F. (2019). Speed-accuracy tradeoffs in brain and behavior: testing the independence of P300 and N400 related processes in behavioral responses to sentence categorization. *Front. Hum. Neurosci.* 13:285. doi: 10.3389/fnhum.2019.00285
- Allen, C. L. (1995). *Case Marking and Reanalysis: Grammatical Relations From Old to Early Modern English*. Oxford: Oxford University Press.
- Allen, M., Poggiali, D., Whitaker, K., Marshall, T. R., and Kievit, R. A. (2019). Raincloud plots: a multi-platform tool for robust data visualization. *Wellcome Open Research* 4:63. doi: 10.12688/wellcomeopenres.15191.1
- Aust, F., and Barth, M. (2018). *Papaja: Create APA Manuscripts with R Markdown*. Available online at: <https://github.com/crsh/papaja>
- Barðdal, J. (2001). The perplexity of Dat-Nom verbs in Icelandic. *Nordic J. Linguist.* 24, 47–70. doi: 10.1080/03325860117730
- Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01
- Bhaskararao, P., and Subbarao, K. V. (eds.). (2004). *Non-Nominative Subjects*, Vol. 1. Amsterdam, Philadelphia, PA: John Benjamins Publishing.
- Bornkessel, I., McElree, B., Schlesewsky, M., and Friederici, A. D. (2004). Multi-dimensional contributions to garden path strength: dissociating phrase structure from case marking. *J. Mem. Lang.* 51, 495–522. doi: 10.1016/j.jml.2004.06.011
- Bornkessel, I., and Schlesewsky, M. (2006a). The extended argument dependency model: a neurocognitive approach to sentence comprehension across languages. *Psychol. Rev.* 113, 787–821. doi: 10.1037/0033-295X.113.4.787
- Bornkessel, I., and Schlesewsky, M. (2006b). The role of contrast in the local licensing of scrambling in German: evidence from online comprehension. *J. Germanic Linguist.* 18, 1–43. doi: 10.1017/S1470542706000018
- Bornkessel, I., Schlesewsky, M., and Friederici, A. D. (2003). Eliciting thematic reanalysis effects: The role of syntax-independent information during parsing. *Lang. Cogn. Process.* 18, 268–298. doi: 10.1080/01690960244000018
- Bornkessel-Schlesewsky, I., and Schlesewsky, M. (2019). Towards a neurobiologically plausible model of language-related, negative event-related potentials. *Front. Psychol.* 10:298. doi: 10.3389/fpsyg.2019.00298
- Bornkessel-Schlesewsky, I., and Schlesewsky, M. (in press). “Cross-linguistic neuroscience of language,” in *The Cognitive Neurosciences, 6th Edn*, eds M. S. Gazzaniga, G. R. Mangun, and D. Poeppel (Cambridge, MA: MIT Press).
- Bornkessel-Schlesewsky, I., Schlesewsky, M., Small, S. L., and Rauschecker, J. P. (2015). Neurobiological roots of language in primate audition: Common computational properties. *Trends Cogn. Sci.* 19, 1–9. doi: 10.1016/j.tics.2014.12.008

FUNDING

Data acquisition for the research reported here was funded by the Max Planck Society for the Advancement of Science and IB-S and DR were at the Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany, when the study was conducted. IB-S acknowledges the support of an Australian Research Council Future Fellowship (FT160100437).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2019.03013/full#supplementary-material>

- Brilmayer, I., Werner, A., Primus, B., Bornkessel-Schlesewsky, I., and Schlesewsky, M. (2019). The exceptional nature of the first person in natural story processing and the transfer of egocentricity. *Lang. Cogn. Neurosci.* 34, 411–427. doi: 10.1080/23273798.2018.1542501
- Bybee, J. (2010). *Language, Usage and Cognition*. Cambridge: Cambridge University Press.
- Choudhary, K. K., Schlesewsky, M., Roehm, D., and Bornkessel-Schlesewsky, I. (2009). The N400 as a correlate of interpretively-relevant linguistic rules: evidence from Hindi. *Neuropsychologia* 47, 3012–3022. doi: 10.1016/j.neuropsychologia.2009.05.009
- Christiansen, M. H., Chater, N., and Culicover, P. W. (2016). *Creating Language: Integrating Evolution, Acquisition, and Processing*. Cambridge, MA: MIT Press.
- Coulson, S., King, J. W., and Kutas, M. (1998). ERPs and Domain Specificity: beating a Straw Horse. *Lang. Cogn. Process.* 13, 653–672.
- Demiral, S. B., Schlesewsky, M., and Bornkessel-Schlesewsky, I. (2008). On the universality of language comprehension strategies: evidence from Turkish. *Cognition* 106, 484–500. doi: 10.1016/j.cognition.2007.01.008
- Eythórsson, T. (2000). “Dative vs. Nominative: changes in quirky subjects in Icelandic,” in *Leeds Working Papers in Linguistics and Phonetics*, Vol. 8, 27–44.
- Faarlund, J. T. (2001). “From ancient Germanic to modern Germanic languages,” in *Language Typology and Language Universals, Vol. II*, eds M. Haspelmath, E. König, W. Oesterreicher, and W. Raible (Berlin: Mouton de Gruyter), 1706–1719.
- Feldman, H., and Friston, K. J. (2010). Attention, uncertainty, and free-energy. *Front. Hum. Neurosci.* 4:215. doi: 10.3389/fnhum.2010.00215
- Fox, J., and Weisberg, S. (2011). *An R Companion to Applied Regression, 2nd Edn*. Thousand Oaks CA: Sage.
- Frisch, S., and Schlesewsky, M. (2001). The N400 indicates problems of thematic hierarchizing. *Neuroreport* 12, 3391–3394. doi: 10.1097/00001756-200110290-00048
- Friston, K. (2005). A theory of cortical responses. *Philos. Trans. R. Soc. B Biol. Sci.* 360, 815–836. doi: 10.1098/rstb.2005.1622
- Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11, 127–138. doi: 10.1038/nrn2787
- Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., et al. (2013). MEG and EEG data analysis with MNE-Python. *Front. Neurosci.* 7:267. doi: 10.3389/fnins.2013.00267
- Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., et al. (2014). MNE software for processing MEG and EEG data. *Neuroimage* 86, 446–460. doi: 10.1016/j.neuroimage.2013.10.027
- Harrington, J. (2012). “The coarticulatory basis of diachronic high back vowel fronting,” *The Initiation of Sound Change: Perception, Production and Social Factors*, eds D. Recasens and M.-J. Solé (Amsterdam, Philadelphia, PA: John Benjamins), 103–122.
- Haupt, F. S., Schlesewsky, M., Roehm, D., Friederici, A. D., and Bornkessel-Schlesewsky, I. (2008). The status of subject-object reanalyses in

- the language comprehension architecture. *J. Mem. Lang.* 59, 54–96. doi: 10.1016/j.jml.2008.02.003
- Hopper, P. J., and Traugott, E. C. (1993). *Grammaticalization*. Cambridge: Cambridge University Press.
- Hörberg, T., Koptjevskaja-Tamm, M., and Kallioinen, P. (2013). The neurophysiological correlate to grammatical function reanalysis in Swedish. *Lang. Cogn. Process.* 28, 388–416. doi: 10.1080/01690965.2011.651345
- Jónsson, J. G. (2003). “Not so quirky: on subject case in Icelandic,” in *New Perspectives in Case Theory*, eds E. Brandner and H. Zinsmeister (Stanford, CA: CSLI Publications), 129–165.
- Jónsson, J. G., and Eythórsson, T. (2005). Variation in subject case marking in Insular Scandinavian. *Nordic J. Linguist.* 28, 223–245. doi: 10.1017/S0332586505001435
- Lenth, R. (2019). *Emmeans: Estimated Marginal Means, aka Least-Squares Means*. Available online at: <https://CRAN.R-project.org/package=emmeans>
- Mueller, J. L., Hahne, A., Fujii, Y., and Friederici, A. D. (2005). Native and nonnative speakers’ processing of a miniature version of Japanese as revealed by ERPs. *J. Cogn. Neurosci.* 17, 1229–1244. doi: 10.1162/0898929055002463
- Ohala, J. J. (1981). “The listener as a source of sound change,” in *Papers from the Parasession on Language and Behavior*, eds C. S. Masek, R. A. Hendrick, M. F. Miller (Chicago, IL: Chicago Linguistic Society), 178–203.
- Platzack, C. (2002). “Nordic language history and generative transformational grammar,” in *The Nordic Languages. An International Handbook of the History of the North Germanic Languages, Vol. 1*, eds O. Bandle, K. Braunmüller, E. H. Jahr, A. Karker, H.-P. Naumann, and U. Telemann (Berlin; New York, NY: Walter de Gruyter), 166–171.
- R Core Team (2018). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- 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
- Schad, D. J., Vasishth, S., Hohenstein, S., and Kliegl, R. (2020). How to capitalize on a priori contrasts in linear (mixed) models: a tutorial. *J. Mem. Lang.* 110:104038. doi: 10.1016/j.jml.2019.104038
- Smith, H. (1994). “Dative Sickness” in Germanic. *Nat. Lang. Linguist. Theor.* 12, 675–736.
- Thráinsson, H. (2014). *On Complementation in Icelandic*. London: Routledge.
- Todd, J., Heathcote, A., Mullens, D., Whitson, L. R., Provost, A., and Winkler, I. (2014). What controls gain in gain control? Mismatch negativity (MMN), priors and system biases. *Brain Topogr.* 27, 578–589. doi: 10.1007/s10548-013-0344-4
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., et al. (2019). Welcome to the tidyverse. *J. Open Source Softw.* 4:1686. doi: 10.21105/joss.01686
- Wilke, C. O. (2019). *Cowplot: Streamlined Plot Theme and Plot Annotations for ggplot2*. Available online at: <https://CRAN.R-project.org/package=cowplot>
- Zaenen, A., Maling, J., and Thráinsson, H. (1990). “Case and grammatical functions: the Icelandic passive,” in *Modern Icelandic Syntax*, eds J. Maling and A. Zaenen (San Diego, CA: Academic Press), 95–136.
- Zhu, H. (2019). *KableExtra: Construct Complex Table with 'kable' and Pipe Syntax*. Available online at: <https://CRAN.R-project.org/package=kableExtra>

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Bornkessel-Schlesewsky, Roehm, Mailhammer and Schlesewsky. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



An Eye-Tracking Study of Sketch Processing: Evidence From Russian

Tatiana E. Petrova^{1*}, Elena I. Riekhakaynen^{2*} and Valentina S. Bratash³

¹ Laboratory for Cognitive Studies, Saint-Petersburg State University, Saint Petersburg, Russia, ² Department of General Linguistics, Saint-Petersburg State University, Saint Petersburg, Russia, ³ Department of Education, Saint-Petersburg State University, Saint Petersburg, Russia

This study investigates the online process of reading and analyzing of sketchnotes (visual notes containing a handwritten text and drawings) on Russian language material. Using the eye-tracking method, we compared the processing of different types of sketchnotes [“path” (trajectory), linear, and radial] and the processing of a verbal text. Biographies of Russian writers were used as the material. In a preliminary experiment, we asked 89 college students to read the biographies and to evaluate each text or sketch using five scales (from –2 to +2). The best example for each of three formats of sketchnotes and a verbal text was chosen. In the main experiment, 21 secondary school students examined four different biographies in four different formats (three sketchnotes and a verbal text), answered to the factual and analytical questions to these texts and estimated the difficulty of each text. We measured the total dwell time, the total fixation count, the average fixation duration for each stimulus as well as for separate zones inside the sketches including verbal and non-verbal information. Our results show that readers process the information better and faster while reading sketchnotes than a verbal text. In the trajectory sketchnotes, the readers followed the order of elements aimed by the author of the sketchnotes better than in the radial and linear sketchnotes. The analysis of participants’ eye movements while processing the stimuli made it possible to propose several recommendations for creating effective sketchnotes.

OPEN ACCESS

Edited by:

Andriy Myachykov,
Northumbria University,
United Kingdom

Reviewed by:

Mikhail Pokhoday,
National Research University Higher
School of Economics, Russia
Marion Krause,
University of Hamburg, Germany

*Correspondence:

Tatiana E. Petrova
t.e.petrova@spbu.ru;
tatianapetrova4386@gmail.com
Elena I. Riekhakaynen
e.riehakajnen@spbu.ru

Specialty section:

This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Psychology

Received: 01 April 2019

Accepted: 07 February 2020

Published: 02 March 2020

Citation:

Petrova TE, Riekhakaynen EI and
Bratash VS (2020) An Eye-Tracking
Study of Sketch Processing: Evidence
From Russian.
Front. Psychol. 11:297.
doi: 10.3389/fpsyg.2020.00297

Keywords: eye-tracking, sketchnoting, text comprehension, text processing, Russian

INTRODUCTION

Nowadays, there is a growing trend toward the use of visual information in various spheres of life (psychology, education, marketing, etc.). Texts containing two non-homogeneous parts – verbal and non-verbal semiotic resources (or “modes”) – have become an integral part of communication. The studies of infographics (graphic visual representation of information), sketchnoting (visual notes including a handwritten text and drawings), advertising copies, multimedia courses integrating the verbal and non-verbal elements are of particular relevance.

The polycode text analysis is traditionally based on the Dual Cording Theory (DCT) (Paivio, 1971, 1986). The theory assumes that there are two distinct cognitive systems: one for processing verbal units and the other one (imagery) for dealing with non-verbal objects/events. Paivio (2006) indicates that the information is represented in the memory by a text and a corresponding illustration, not just by a text. It is assumed that the information in a polycode text is double-decoded: the concept of an image is “superimposed” on the concept of a verbal text, the interaction of these two concepts leads to the creation of a general concept (meaning) of the text (Telminov, 2009; Fernández-Fontecha et al., 2018). Independent parts of a polycode text interact and create a “holistic experience,” the combination of the visual language with the written language.

In many studies, the influence of visual components on the comprehension of the whole polycode text is evaluated by offline tests (questionnaires, scales, etc.) (Cohn, 2016). However, these methods can only measure the result of the comprehension process. Thus, the identification of the particular elements which influence different stages of the process is difficult. Eye tracking techniques provide online information about learners' behavior during text reading. As Rayner (1998) points out, by using eye tracking, one can study reading as a process, instead of "a mere end-result." During the last 20 years, a lot of empirical and experimental evidence on online processing of polycode texts (including comics and visual narratives) appeared. One of the paradigms is called Visual Language Theory (VLT) which describes how visual lexical items are read taking into consideration the structure of polycode samples and trying to develop the "narrative grammar of sequential images" (Cohn, 2018). This approach argues that verbal and non-verbal components operate in parallel as interfering structures.

Polycode texts are regarded as a new type of texts used in education (Kazakova, 2016). They have become a crucial part of teaching in a wide range of academic and practical disciplines (Altieri, 2017; Chandler, 2017). The educational aspects of the polycode text processing are considered among others in the Cognitive Theory of Multimedia Learning (CTML) (Mayer, 2009). This theory assumes that the combination of verbal information and pictures makes it easier for learners to understand and memorize a text. While studying the processing of verbal and non-verbal information, Levie and Lentz (1982) concluded that the information supported by both a text and a picture is acquired much better. The more switches there are between a text and an image while reading a polycode text, the better a reader understands the material (e.g. Mason et al., 2013; Scheiter and Eitel, 2015). It has been shown that if students do not pay enough attention to the pictures and focus mainly on the text zones or untimely correlate verbal and visual information, the effectiveness of training falls significantly (Hannus and Hyönä, 1999; Schwonke et al., 2009; Schmidt-Weigand et al., 2010; Cromley et al., 2013a,b; Mason et al., 2013; Renkl and Scheiter, 2017). Moreno and Mayer (2002) and Johnson and Mayer (2012) tried to solve this problem by means of additional instructions. Ozcelik et al. (2009) and Scheiter and Eitel (2015) used spatial-color schemes reducing the distance between the text and the picture or highlighting the corresponding verbal and non-verbal elements in one color. However, such tools were shown to be effective only for poorly prepared students, but do not influence the results of students with a higher level of training (Kalyuga et al., 2003; Sweller et al., 2003; Kalyuga, 2007; Richter et al., 2017).

This research is conducted within the CTML and is aimed to study the processing of sketchnoting (or visual notes) as an example of a polycode (multimodal) text. As far as we know, sketchnotes have never become the object of a psycholinguistic research using online methods, although they seem to be worthwhile both for learning how we process multimodal information and for educational purposes as a new type of data compression. Sketchnoting combines various ways of presenting information and includes such uncommon for other types of polycode texts elements as hand-drawn typography, handwritten

(not printed) texts, and many different visual components: drawings, arrows, lines, and dots (Rohde, 2013). Moreover, there are several distinct types of the organization of the material in sketchnotes. There are seven types of sketch structures: (1) path (trajectory; with arrows helping to navigate the text), (2) linear (information and visual components are arranged as in a normal verbal text – lines going from the left to the right), (3) radial (the main idea of the text is in the middle of the list surrounded by other text elements), (4) vertical (text elements are organized vertically: from the top to the bottom of the page), (5) modular (each piece of information forms a separate block), (6) skyscrapers (the information is organized in several vertically stretched rectangles), and (7) popcorn (with random arrangement of all blocks of information) (Rohde, 2013: 90) (see the layouts of all sketch structures in **Supplementary Figure 1**). Thus, we can compare how different structures are processed, explore the impact of a sketch type on the navigation decisions, and find out the most efficient sketch structure for retrieving the information. The aim of our study was to compare the processing of sketchnotes and a verbal text and to choose the best type of sketchnotes for transferring the information to a reader.

PRELIMINARY EXPERIMENT

Goal

The goal of the preliminary experiment was to choose the stimuli for the main experiment, i.e. the sketchnotes of different structures and a verbal text that are evaluated as the most attractive (interesting, informative, good-structured, etc.) by school children – native speakers of Russian.

Material

We chose the biographies of four Russian poets for our study. A biography is usually a stereotyped text with a standard structure (including such common information as the years of life, the place of birth, education, some information about the family, profession, interests, the main stages of life, etc.). Biographies are often used while studying literature at school.

The poets were as follows: O. Mandelshtam, M. Voloshin, Z. Gippius, and I. Severyanin. All of them lived in the first half of the 20th century and their poems are not included in the obligatory school program in Russia. Thus, we can assume that the background of our participants who were school children did not influence significantly their performance in the experiments as most probably they were not familiar with the biographies we had chosen for our study.

The initial biographies were in verbal format (plain texts) and taken from one and the same resource (guide on literature). All the texts were of the same size and comparable level of readability¹ (**Supplementary Table 1**). To get the stimuli, we converted all biographies into three main sketch formats (that contain the features of all other types of sketchnotes): trajectory, linear, and radial using the guidelines provided in Rohde (2013). Thus, the material of the experiment consisted of 4 different

¹Checked via <http://readability.io/>.

verbal texts and 12 sketchnotes – three formats for each of four different biographies². The readability level in all types of the sketchnotes was lower than in the plain texts.

Procedure and Participants

We asked 89 Russian school children (45 girls) to read the biographies and to evaluate each text or sketch using five scales (from -2 to $+2$ each): non-informative – informative, difficult to understand – easy to understand, not interesting – interesting, difficult to retell – easy to retell, bad structure – good structure. We used the Latin Square design. Every participant read four different biographies each of them presented either as a verbal text or in one of three sketch formats. Thus, every participant saw each type of the text and each biography only once. All the stimuli were presented in randomized order. The experiment lasted around 20 min for each participant.

Results

For each of 16 stimuli, we summed up the scores from all five scales for each participant and compared these aggregate scores for different formats of presentation of one and the same biography (using ANOVA and the Kruskal–Wallis test for independent samples according to the type of data distribution). We found the factor of the format of presentation to be significant for three out of four biographies. To reveal the best format for each of these three biographies we performed the unpaired two-sample *t*-tests for each pair of formats within each biography. Surprisingly, for all four biographies, we got quite high aggregate scores for the verbal format. The sketchnotes that turned out to be significantly different from the verbal text are marked with an asterisk (*) on **Supplementary Figures 2–5**. We did not find the significant difference between the verbal text and the trajectory sketchnotes for any of the biographies.

For the main experiment, we had to choose four different formats of presentation from the preliminary experiment (linear sketchnotes, radial sketchnotes, trajectory sketchnotes, and a verbal text). As we planned to show all four formats for every participant in the main experiment, we could choose only one stimulus for each biography. Thus, taking into consideration this condition, we were choosing among the stimuli with the highest aggregate scores for each biography and finally got the following set of stimuli: (1) the biography of Z. Gippius – the verbal text; (2) the biography of I. Severyanin – the trajectory sketchnotes; (3) the biography of O. Mandelshtam – the radial sketchnotes; and (4) the biography of M. Voloshin – the linear sketchnotes. The text parts of the sketchnotes 2–4 were of a comparable readability level (**Supplementary Table 2**) and had the equal number of pictures.

MAIN EXPERIMENT

Hypothesis

The hypothesis of the experiment was that readers process different text formats differently, trajectory sketchnotes being

easier to process and understand than other types of sketchnotes and a verbal text.

Participants

Twenty-one native speakers of Russian (secondary school children, 13–18 years old, 11 girls), who had not participated in the preliminary experiment, took part in the main experiment on voluntary basis. All subjects had normal or corrected to normal vision.

Procedure

We conducted an eye-tracking experiment. We used a SR EyeLink 1000 plus eye tracker (SR Research Ltd., ON, Canada) with a head holder (“desktop mode” configuration) and 27” LCD monitor (Acer v276hl) with a refresh rate of 60 Hz (screen resolution 1920 × 1080) to record the eye-movements of the participants. Viewing distance was 87 cm. It differs from the recommended eye-to-monitor distance for EyeLink 1000+, but it was the only way to place the monitor in the given conditions. We conducted several pilot trials and revealed that a participant could see all the letters and pictures of the stimuli at this distance and the nine-point calibration and validation were successful. The average error level during calibration was $<0.5^\circ$; the threshold was 1° . Although viewing was binocular, we recorded participants’ dominant eye. All but two of the participants had the right dominant eye. We used SR Research Experiment Builder to create and run the experiment and EyeLink Data Viewer to analyze the results.

After successful calibration and validation each subject received an instruction to examine four different biographies sequentially presented on the computer screen and be ready to answer the questions after each text or sketch. All biographies were presented on the computer screen for 5 min. The participants were free to press the spacebar button if they were ready to answer the questions earlier than after 5 min. For each biography, we prepared four factual questions, three questions revealing the general comprehension of the sketch or the text and one rating scale question for estimating whether the text was difficult or easy to understand (from -2 = very easy to $+2$ = very difficult). The list of questions for each sketch and a verbal text can be found here: <https://drive.google.com/drive/folders/1xgpKcymbzI28bYoy3QpGINvoQDcQzcGz?usp=sharing>.

The participants answered orally. One of the experimenters marked correct answers in a special paper form. We also used a digital voice recorder Olympus WS-650S to record the participants’ responses to be able to revise the data. We used drift correction before presenting each text or sketch and if it turned out to be unsuccessful, we performed recalibration. The experiment lasted for about 40 min (including the calibration and recalibration period).

The experiment was conducted *in July 2018* at the Educational Centre “Sirius” (Sochi, Russia) in accordance with the Declaration of Helsinki and the existing Russian and international regulations concerning ethics in research. It was approved by the Ethics Committee of Saint-Petersburg State University in June 2018. As the participants were under

²The data are available at: <https://drive.google.com/drive/folders/1xgpKcymbzI28bYoy3QpGINvoQDcQzcGz?usp=sharing>

18 years old, we obtained written informed consents for their participation in the experiment from their parents.

Measures

We considered several global eye movement measures traditionally used for studying polycode text processing (dwell time, total fixation count, average fixation duration). As the aim of the research was to compare the processing of different types of the sketchnotes, we also calculated the number of deviations from the trajectory aimed by the author of the sketchnotes while each participant processed every sketch. We also segmented all sketchnotes into interest areas, i.e. verbal and non-verbal elements of the sketch, and analyzed interest area dwell time, interest area first run dwell time, interest area fixation count for each verbal and non-verbal zone of the texts in order to compare the processing of different structural elements of the sketchnotes. The number of correct answers to the factual and analytical questions and the subjective difficulty of different stimuli were also analyzed.

Results

Due to some technical problems, we did not manage to record the eye-movements of three participants while processing one of the formats (twice the verbal text and once the trajectory sketchnotes) and the eye-movements of one more participant while processing two formats (the verbal text and the linear sketchnotes). Thus, when we compared the processing of different formats by one and the same participant, we excluded the results of these four participants.

The Friedman test showed the influence of the factor “Format type” on the parameters “Dwell time” [$\chi^2(3) = 19.24, p < 0.001$], “Total fixation count” [$\chi^2(3) = 23.61, p < 0.001$], and “Average fixation duration” [$\chi^2(3) = 12.88, p = 0.005$]. According to Conover’s *post hoc* tests, sketchnotes of any format were read significantly more quickly and with a smaller number of fixations than the text whereas the processing of different types of sketchnotes did not differ significantly (see **Supplementary Tables 3, 4** and **Supplementary Figures 6, 7**, respectively). The difference in the average fixation duration is not that clear-cut. There is no significant difference between the average fixation duration for the trajectory and radial sketchnotes ($p = 0.882$), the linear and radial sketchnotes ($p = 0.059$), and the linear sketchnotes and the text ($p = 0.186$) whereas in all other pairs we did find significant differences. The mean fixation duration for the text is shorter than for any type of the sketchnotes, but the results not of all the participants follow this tendency.

While reading the trajectory sketchnotes the participants significantly more often ($p = 0.019$ in the Binomial test) followed the order of reading the sketch elements aimed by the author of the sketchnotes than diverged from it (we considered that the participant diverged from the aimed trajectory if there were three or more deviations) (**Supplementary Table 5**). While processing the radial sketchnotes, there were more participants who followed the order of reading than those who did not, but the difference was not statistically significant ($p = 0.245$). Only around 30% of the participants followed the order of the sketched elements

aimed by the author while processing the linear sketchnotes (**Supplementary Table 5**).

Our results also revealed that all the sketchnotes were subjectively evaluated by the participants to be easier to understand than the verbal text (the median value $Me = 2$ and $Me = 1$, respectively). The participants answered correctly to significantly more questions after all sketchnotes than after the verbal text. The influence of this factor was shown by the Friedman test: $\chi^2(3) = 18.26, p < 0.001$; the Conover’s *post hoc* tests demonstrated the significant difference between the results for the text and for all types of the sketchnotes (**Supplementary Table 6**). The same was true if we compared the number of correct answers only to factual questions. For analytical questions, we got significantly better results for the linear and trajectory sketch, whereas for the radial sketch the distribution of correct and incorrect answers did not show statistically significant difference from the results for the verbal text (**Supplementary Table 7**).

The radial sketch (the biography of Mandelshtam; see **Supplementary Figure 8**) is of particular interest since it contains both horizontal and diagonal zones. We compared the processing of a horizontal zone (interest areas “Mtext_1_mood” and “Mtext_2_Pushkin” together) to the processing of a diagonal zone (“Mtext_6_epigramma_diagonal”) of the same size (containing equal number of symbols: 154 and 156, respectively) and revealed that the dwell time for the horizontal zone was significantly less than for the diagonal zone ($W = 45, p = 0.024$).

We compared the processing of zones containing verbal and non-verbal information in the linear sketch (the biography of Voloshin; see **Supplementary Figure 9**) as it was the only sketchnotes where there were several comparable pieces of information presented both in verbal and non-verbal format. These were the portraits of Russian poets and writers (“Bimage_2_Cvetaeva,” “Bimage_13_Beli,” “Bimage_14_Gorki”) and signs with their names (“Btext_12_Cvetaeva,” “Btext_13_Beli,” “Btext_14_Gorkij”). These zones of interest were of the same size and the same content. We revealed that the verbal components in all three image-text pairs were processed less quickly (Tsvetaeva – $W = 176, p = 0.006$; Belyj – $187, p < 0.001$; Gor’kij – $156, p = 0.015$). However, we didn’t find this effect for the portrait of the main hero. There was no significant difference between the processing of the portrait of Voloshin (“Bimage_1_partrait”) and the verbal zone with his name (“Btext_1_titel”) above it ($W = 120, p = 0.596$).

The average time spent on the title zones turned out to differ significantly in all three sketchnotes being the longest for the biography of Mandelshtam (the radial sketchnotes) and the shortest for the biography of Severyanin (the trajectory sketchnotes) [see **Supplementary Figure 11**, **Supplementary Table 8**, and the heat maps (**Supplementary Figures 8–10**)].

DISCUSSION

In our study, we found that the processing of any type of sketchnotes where verbal information is combined with non-verbal differs significantly from the processing of the verbal text.

These results correlate with the previous studies that showed that an image and a written text presented together can contribute to a better understanding of the information than if they are presented separately (Schnotz, 2005) and with the CTML (Mayer, 2009) that assumes that a multimodal text is an effective form of learning as it implicates switching of attention between a text and an image and establishes links between the two elements. The so-called multimedia effect helps to integrate the new information in the cognitive system and to remember it.

As it was shown in numerous eye-tracking studies, a text is read according to the *F*-shaped scanning pattern which is characterized by many fixations concentrated at the top-left part of the screen (Pernice, 2017). We got the same results for the processing of the biography presented as a verbal text. There were more gazes on the first lines than on the subsequent ones. The first several words on the left of each line received more fixations than subsequent words in the same line (**Supplementary Figure 12**). For all the sketchnotes we analyzed, the reading patterns were usually text-directed. This result correlates with other studies of polycode texts that showed that the text zones receive more attention than the picture zones (Rayner et al., 2001; Petrova and Riekhakaynen, 2019). Lee and Wu (2017) have also shown that a picture or a geometric figure attracts less reader's attention than a text in the process of scanning math texts. Although the sketchnotes we analyzed represented three different types of information organization, we did not find any significant differences in the time of their processing and, the number of fixations, subjective evaluation, and the number of correct answers to the after-the-text questions. However, while processing the trajectory sketchnotes the participants followed the order of reading aimed by the author better than while processing the linear and radial sketchnotes. We presume that, despite the fact that the participants did not pay much attention to the zones with small arrows that were numerous in the trajectory sketch (there were few fixations on them), these arrows helped not to deviate. We also did find some differences in the average fixation duration between the sketchnotes we analyzed. These results require further consideration, but we presume that the factors influencing the average fixation duration include the font size, the number of pictures in the texts, as well as the individual strategies of participants.

Our results also allow to discuss some basic principles of the polycode text structure. Although the pictures usually attract less attention than the verbal text containing the same information, the portraits of the main characters are normally scanned more attentively than other images. This finding is close to some recent face recognition eye-tracking studies and recommendations to use portraits and pictures of a person's face in order to increase reader's attention to a website (Patel, 2014) and banner advertisements (Sajjacholapunt and Ball, 2014). At the same time, the results we got on how a reader scan the titles of the sketchnotes do not correspond to the prior studies that showed that readers paid more attention to headings (e.g. Hyönä et al., 2002; Hyönä and Lorch, 2004; Lemarié et al., 2012) and found them useful when reading a text (Hartley and Trueman, 1985; Yussen et al., 1993), encoding the topic-comment structure of a text and recalling the text content (Lorch and Lorch, 1995).

It was revealed that different types of headings influence the process of searching the text and the sequence of examination of text elements (Klusewitz and Lorch, 2000). Our results show that the participants do not pay much attention to the title zones. However, we still believe that the headings are helpful for finding the target information in the text and can be used to guide the process of examining the text or sketch. According to the results we received, to attract more attention the title in a polycode text should be somehow included in the overall structure of the sketchnotes or should be placed in non-standard way.

CONCLUSION

Reading is a complex task that depends on many different cognitive processes. Numerous experiments have shown that text understanding is a complex multistep process. The comprehension of a written text includes – among others – the recognition and pattern analysis of letters, graphics, and structural components. Recent cognitive-orientated research shows that the text type is among the readability categories. The aim of the present study was to reveal whether a sketch or a verbal text is easier to process and better to use for retrieving the essential information.

Eye-tracking studies of the processing of Russian texts are not numerous. They are mainly focused on the recognition of a regular verbal text (Bezrukikh and Ivanov, 2013, 2014, 2015; Kornev et al., 2014; Petrova, 2016; Korneev et al., 2017a,b). There is only one eye-tracking research on Russian (Petrova and Riekhakaynen, 2019) in which the processing of a polycode text, namely infographics, has been studied. It was one of the first steps to reveal how readers integrate text-figure information when reading and understanding infographics.

The results of the present study have shown that a sketch of any format is read faster than a verbal text. It is worth mentioning that the percentages of correct answers to the after-the-text questions are normally higher after processing sketchnotes than after reading a verbal text. The trajectory (path) seems to be the most efficient type of sketchnoting because it clearly shows a reader the order of reading aimed by the author of the sketchnotes.

The analysis of participants' eye movements while processing the stimuli allowed us to propose a number of recommendations for creating sketchnotes: (1) diagonal position of the text is not efficient because such zones are read significantly slower than the zones where the text is arranged horizontally; (2) it is better to control the reader's attention with the arrows as they show the order of acquiring the information presumed by the author of a sketch and thus help to learn the text faster; and (3) it is important to duplicate the information from the title somewhere inside the sketchnotes or to integrate the title into the sketch to attract reader's attention to it.

We suppose that visual notes can be a functional alternative of a traditional verbal summary and this format can diversify the educational process. It is possible to recommend using sketchnoting as an alternative way of processing large blocks of information, when a reader can decide himself what type of

summary to choose. The data obtained open perspectives for further investigation of the reading process, means of resolving ambiguity in the different text types, and the relationship between verbal and non-verbal parts of the text.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**.

ETHICS STATEMENT

The experiment was conducted in July 2018 at the Educational Centre “Sirius” (Sochi, Russia) in accordance with the Declaration of Helsinki and the existing Russian and international regulations concerning ethics in research. It was approved by the Ethics Committee of Saint-Petersburg State University in June 2018. As the participants were under 18 years old, we obtained written informed consents for their participation in the experiment from their parents.

AUTHOR CONTRIBUTIONS

TP: main idea, data collection for the experiments, introduction, and discussion. ER: the eye-tracking experiment and analysis of

the results, figures, and tables. VB: choosing the stimuli, data collection, and creating sketchnotes. All authors contributed to the research and to the manuscript, and agreed to be accountable for the content of the work.

FUNDING

This study was supported by the research grant no. 18-00-00640 “Linguistic information processing under ambiguity: activation and competition of variants” from the Russian Foundation for Basic Research.

ACKNOWLEDGMENTS

We thank our students Elizaveta Zhukatinskaya, Anastasija Salnikova, Ekaterina Zhelezova, Marina Solnceva, Anastasija Suhareva, Alexandra Bervinova for the help with collecting and analyzing the data. We also thank the Educational Centre “Sirius” (Sochi, Russia) for providing the field for the research and allowing access to their equipment and students.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2020.00297/full#supplementary-material>

REFERENCES

- Altieri, J. L. (2017). From sketchnotes to think-alouds: addressing the challenges of social studies text. *Soc. Stud. Young Learn.* 30, 8–12.
- Bezrukikh, M. M., and Ivanov, V. V. (2013). Eye-movements while reading as an indicator of skill acquisition. *Fiziol. Cheloveka* 39, 83–93. doi: 10.7868/S0131164612060045
- Bezrukikh, M. M., and Ivanov, V. V. (2014). Oculomotor activity while reading in children with different reading skills. (Message 1). Oculomotor activity in 6–7-year-old children with good and poor reading skills. *Novye Issledovaniya* 41, 67–76.
- Bezrukikh, M. M., and Ivanov, V. V. (2015). Reading oculomotor activity in children with different reading skills. (Message 2). Peculiarities of oculomotor activity in 9–10-year-old children with good and poor reading skills. *Novye Issledovaniya* 43, 4–12.
- Chandler, C. (2017). Improving student note-taking skills. *Educ. Dig.* 82, 54–56.
- Cohn, N. (2016). A multimodal parallel architecture: a cognitive framework for multimodal interactions. *Cognition* 146, 304–323. doi: 10.1016/j.cognition.2015.10.007
- Cohn, N. (2018). “Visual language theory and the scientific study of comics,” in *Empirical Comics Research: Digital, Multimodal, and Cognitive Methods*, eds A. Dunst, J. Laubrock, and J. Wildfeuer (London: Routledge), 305–328. doi: 10.4324/9781315185354-15
- Cromley, J. G., Bergey, B. W., Fitzhugh, S. L., Newcombe, N., Wills, T. W., Shipley, T. F., et al. (2013a). Effectiveness of student-constructed diagrams and self-explanation instruction. *Learn. Instr.* 26, 45–58. doi: 10.1016/j.learninstruc.2013.01.003
- Cromley, J. G., Perez, A. C., Fitzhugh, S., Newcombe, N., Wills, T. W., and Tanaka, J. C. (2013b). Improving students’ diagrammatic reasoning: a classroom intervention study. *J. Exp. Educ.* 81, 511–537. doi: 10.1080/00220973.2012.745465
- Fernández-Fontecha A., O’Halloran, K. L., Tan, S., and Wignell, P. A. (2018). Multimodal approach to visual thinking: the scientific sketchnote. *Vis. Commun.* 18, 5–29. doi: 10.1177/1470357218759808
- Hannus, M., and Hyönä, J. (1999). Utilization of illustrations during learning of science textbook passages among low- and high-ability children. *Contemp. Educ. Psychol.* 24, 95–123. doi: 10.1006/ceps.1998.0987
- Hartley, J., and Trueman, M. (1985). A research strategy for text designers: the role of headings. *Instr. Sci.* 14, 99–155. doi: 10.1007/BF00052394
- Hyönä, J., and Lorch, R. (2004). Effects of topic headings on text processing: evidence from adult readers’ eye fixation patterns. *Learn. Instr.* 14, 131–152. doi: 10.1016/j.learninstruc.2004.01.001
- Hyönä, J., Lorch, R. F. Jr., and Kaakinen, J. K. (2002). Individual differences in reading to summarize expository text: evidence from eye fixation patterns. *J. Educ. Psychol.* 94, 44–55. doi: 10.1037/0022-0663.94.1.44
- Johnson, C. I., and Mayer, R. E. (2012). An eye movement analysis of the spatial contiguity effect in multimedia learning. *J. Exp. Psychol. Appl.* 18, 178–191. doi: 10.1037/a0026923
- Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educ. Psychol. Rev.* 19, 509–539. doi: 10.1007/s10648-007-9054-3
- Kalyuga, S., Ayres, P., Chandler, P., and Sweller, J. (2003). The expertise reversal effect. *Educ. Psychol.* 38, 23–31. doi: 10.1207/S15326985EP3801_4
- Kazakova, E. I. (2016). Texts of the new nature. *Psychol. Sci. Educ.* 21, 102–109.
- Klusewitz, M. A., and Lorch, R. F. (2000). Effects of headings and familiarity with a text on strategies for searching a text. *Mem. Cognit.* 28, 667–676. doi: 10.3758/BF03201256
- Korneev, A. A., Akhutina, T. V., and Matveeva, E. Y. (2017a). “Analysis of elementary school children’ eye movements while reading,” in *Cognitive Science in Moscow: New Studies*, eds E. V. Petchenkova, and M. V. Falikman (Moscow: Buki Vedi), 158–162.

- Korneev, A. A., Matveeva, E. Y., and Akhutina, T. (2017b). Reading in Russian primary school children: an eye tracking study. *J. High. Sch. Econ.* 14, 219–235. doi: 10.17323/1813-8918.2017.2.219.235
- Kornev, A. N., Balcuniene, I., Selezneva, M. G., and Sorokin, A. A. (2014). “How to compute reading development strategy? Evidence from Russian-speaking children,” in *Human Language Technologies – The Baltic Perspective*, eds A. Utka, G. Grigonyte, J. Kapociute-Dzikiene, and J. Vaicenoniene (Washington, DC: IOS Press), 196–201.
- Lee, W.-K., and Wu, C.-J. (2017). Eye movements in integrating geometric text and figure: scanpaths and given-new effects. *Int. J. Sci. Math. Educ.* 16, 699–714. doi: 10.1007/s10763-016-9790-2
- Lemarié, J., Lorch, R., and Péry-Woodley, M.-P. (2012). Understanding how headings influence text processing. *Discours* 10, 2–22. doi: 10.4000/discours.8600
- Levie, W. H., and Lentz, R. (1982). Effects of text illustrations: a review of research. *Educ. Commun. Technol.* 30, 195–232.
- Lorch, R. F. Jr., and Lorch, E. P. (1995). Effects of organizational signals on text processing strategies. *J. Educ. Psychol.* 87, 537–544. doi: 10.1037/0022-0663.87.4.537
- Mason, L., Tornatora, M. C., and Pluchino, P. (2013). Do fourth graders integrate text and picture in processing and learning from an illustrated science text? Evidence from eye-movement patterns. *Comput. Educ.* 60, 95–109. doi: 10.1016/j.compedu.2012.07.011
- Mayer, R. E. (2009). *Multimedia Learning*, 2nd Edn. Cambridge, MA: Cambridge University Press.
- Moreno, R., and Mayer, R. E. (2002). Verbal redundancy in multimedia learning: when reading helps listening. *J. Educ. Psychol.* 94, 156–163. doi: 10.1037/0022-0663.94.1.156
- Ozcelik, E., Karakus, T., Kursun, E., and Cagiltay, K. (2009). An eye tracking study of how color-coding affects multimedia learning. *Comput. Educ.* 53, 445–453. doi: 10.1016/j.compedu.2009.03.002
- Paivio, A. (1971). *Imagery and Verbal Processes*. New York, NY: Holt Rinehart and Winston.
- Paivio, A. (1986). *Mental Representations: A dual Coding Approach*. Oxford: Oxford University Press.
- Paivio, A. (2006). *Dual Coding Theory and Education. Draft Chapter for the Conference on Pathways to Literacy Achievement for High Poverty Children*. Ann Arbor, MI: The University of Michigan School of Education.
- Patel, N. (2014). *8 Powerful Takeaways from Eye Tracking Studies*. Available at: <https://www.quicksprout.com/2014/04/16/8-powerful-takeaways-from-eye-tracking-studies/> (accessed April 1, 2019).
- Pernice, K. (2017). *F-Shaped Pattern of Reading on the Web: Misunderstood, But Still Relevant (Even on Mobile)*. Available at: <https://www.nngroup.com/articles/f-shaped-pattern-reading-web-content/> (accessed April 1, 2019).
- Petrova, T. (2016). “Eye movements in reading the texts of different functional styles: evidence from Russian,” in *Intelligent Decision Technologies 2016. Smart Innovation, Systems and Technologies*, eds I. Czarnowski, A. Caballero, R. Howlett, and L. Jain (Cham: Springer). 57, 285–298. doi: 10.1007/978-3-319-39627-9
- Petrova, T. E., and Riekhakaynen, E. I. (2019). Processing of verbal and non-verbal patterns: an eye-tracking study of Russian. *Adv. Intell. Syst. Comput.* 797, 269–276. doi: 10.1007/978-981-13-1165-9_24
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychol. Bull.* 124, 372–422. doi: 10.1037/0033-2909.124.3.372
- Rayner, K., Rotello, C., Stewart, A., Keir, J., and Duffy, S. (2001). Integrating text and pictorial information: eye movements when looking at print advertisements. *J. Exp. Psychol. Appl.* 7, 219–226. doi: 10.1037/1076-898x.7.3.219
- Renkl, A., and Scheiter, K. (2017). Studying visual displays: how to instructionally support learning. *Educ. Psychol. Rev.* 29, 599–621. doi: 10.1007/s10648-015-9340-4
- Richter, J., Scheiter, K., and Eitel, A. (2017). Signaling text-picture relations in multimedia learning: the influence of prior knowledge. *J. Educ. Psychol.* 110, 544–560. doi: 10.1037/edu0000220
- Rohde, M. (2013). *The Sketchnote Handbook: The Illustrated Guide to Visual Note Taking*. Berkeley, CA: Peachpit Publishing Company.
- Sajjacholapunt, P., and Ball, L. J. (2014). The influence of banner advertisements on attention and memory: human faces with averted gaze can enhance advertising effectiveness. *Front. Psychol.* 5:166. doi: 10.3389/fpsyg.2014.00166
- Scheiter, K., and Eitel, A. (2015). Signals foster multimedia learning by supporting integration of highlighted text and diagram elements. *Learn. Instr.* 36, 11–26. doi: 10.1016/j.learninstruc.2014.11.002
- Schmidt-Weigand, F., Kohnert, A., and Glowalla, U. (2010). Explaining the modality and contiguity effects: new insights from investigating students’ viewing behaviour. *Appl. Cogn. Psychol.* 24, 226–237. doi: 10.1002/acp.1554
- Schnotz, W. (2005). “An integrated model of text and picture comprehension,” in *The Cambridge Handbook of Multimedia Learning*, ed. R. E. Mayer (Cambridge, MA: Cambridge University Press), 49–69.
- Schwonke, R., Berthold, K., and Renkl, A. (2009). How multiple external representations are used and how they can be made more useful. *Appl. Cogn. Psychol.* 23, 1227–1243. doi: 10.1002/acp.1526
- Sweller, J., Kalyuga, S., Ayres, P., and Chandler, P. (2003). The expertise reversal effect. *Educ. Psychol.* 38, 23–31. doi: 10.1207/S15326985EP3801
- Telminov, G. N. (2009). Advertisements in the internet as a type of creolized texts. *Vestnik Nizhegorodskogo universiteta im. N.I. Lobachevskogo.* 5, 300–304.
- Yussen, S. R., Stright, A. D., and Payne, B. (1993). Where is it? Searching for information in a college textbook. *Contemp. Educ. Psychol.* 18, 240–257. doi: 10.1006/ceps.1993.1019

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Petrova, Riekhakaynen and Bratash. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Advantages of publishing in Frontiers



OPEN ACCESS

Articles are free to read for greatest visibility and readership



FAST PUBLICATION

Around 90 days from submission to decision



HIGH QUALITY PEER-REVIEW

Rigorous, collaborative, and constructive peer-review



TRANSPARENT PEER-REVIEW

Editors and reviewers acknowledged by name on published articles

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org

Contact us: info@frontiersin.org | +41 21 510 17 00



REPRODUCIBILITY OF RESEARCH

Support open data and methods to enhance research reproducibility



DIGITAL PUBLISHING

Articles designed for optimal readership across devices



FOLLOW US

[@frontiersin](https://twitter.com/frontiersin)



IMPACT METRICS

Advanced article metrics track visibility across digital media



EXTENSIVE PROMOTION

Marketing and promotion of impactful research



LOOP RESEARCH NETWORK

Our network increases your article's readership