ISSN 2752-7417

Undergraduate Linguistics Association of Britain

Proceedings of ULAB 2023

University of Manchester

13-15 April 2023

Edited by James Berriman, Eden Elia Müller, & Lydia Wiernik

Undergraduate Linguistics Association of Britain

ULAB XIII Proceedings
University of Manchester
13-15 April 2023

Editors: James Berriman, Eden Elia Müller, Lydia Wiernik
Copyeditors: Nicholas Daines, Alice Duncan, Alice Eddyshaw, Ilias Kolokousis
Cover design: Eden Elia Müller

ISSN 2752-7417

Contents

ULAB National Committee, 2022-23

Nicole Chan, *University of Edinburgh*	Chair
Nicholas Daines, *University of Edinburgh*	Vice Chair
Ro Dhasmana, *University of Aberdeen*	Secretary
Kirsty Gardiner, *University of Aberdeen*	Treasurer
Lydia Wiernik, *University of Edinburgh*	Journal Editor
Kitty Liu, *University of Cambridge*	Magazine Editor-in-Chief
Matthew Hale, *University of Edinburgh*	Social Media Coordinator
Caitlin Wilson, *University of Oxford*	Archivist
Kazune Sato, *University of Cambridge*	Webmaster
Beatrix Livesey-Stephens, *University of Abertay*	Events Coordinator
Finn Amber, *University of Cambridge*	Accessibility Officer
Cliodhna Hughes, *University of Sheffield*	Opportunities Coordinator
Robin Tang, *University of Manchester*	Institutional Representative Coordinator
Nicole Chan, *University of Edinburgh*	Institutional Representatives
Kitty Liu, *University of Cambridge*	
Jade Pathak, *University of Manchester*	
Zsofia Varga, *University College London*	
Naomi Black, *University of Aberdeen*	
Emily Shepherdson, *University of York*	
Becky Hunt, *University of Sussex*	
Eleanor Streatfield, *University of Birmingham*	
Holly Taylor, *University of Warwick*	

ULAB Local Committee, 2022-23

Jade Pathak	Chair
Santiago Chau	Vice Chair
Cassandra Papp	Secretaries
Kizzy Fennell	
Ekin Ozbudkan	Treasurers
Maelys LeFranc	
Hannah Hatfield	Workshop & Panel Coordinator
Sophie Pitchford	Social Secretary/Social Media Coordinator
Lucas Pear	Publicist/Social Media Coordinator
Kate Towlerton	Social and Publicist Support
Elouise Brennan	
Kizzy Fennell	Photography Coordinator

Foreword from the National Secretary, 2022-2023

Ro Dhasmana, University of Aberdeen

The 2023 Conference Proceedings are a snapshot of the undergraduate work presented by young linguists at ULAB XII, organised by students from the University of Manchester and Manchester Metropolitan University. This year marked ULAB's second hybrid conference, with presenters online and in-person coming from all over the world to share their undergraduate research. In addition to individual work, we saw presentations from this year's Research Buddy Scheme, connecting students from different backgrounds with similar interests and encouraging them to carry out a collaborative research project.

Above all, ULAB strives for accessibility and inclusion in undergraduate linguistics. This has been exemplified in many different ways over the year - a variety of in-person and online events, JoULAB's push for unmarked dissertations and solidarity with the UCU's Marking Boycott, increased bursaries for students attending the conference, and a series of schemes such as the Essay Competition and Research Buddies, providing new opportunities for linguists. The Proceedings are an opportunity for undergraduate linguists to have their work published, and for students past and present to be inspired by their research.

Finally - thanks must go to Nicole, our National Chair, and Jade, our Local Chair, for the hard work and dedication they have shown during this busy year. Last but by no means least, a massive shout-out to the JoULAB committee, led by the unstoppable Lydia, for compiling and copyediting the Proceedings - your drive and energy are a credit to ULAB!

Foreword from the Local Chair, 2022-2023

Jade Pathak, University of Manchester

I had the amazing privilege of being the local chair for the ULAB 2023 conference, which marked the first year it had been held in the city of Manchester! We welcomed students from universities across the UK and were able to extend the offering of ULAB to further afield with remote delivery through a hybrid in-person/Zoom conference, following the success of the VirtualULAB 2021 and ULAB Edinburgh 2022. Over three busy days, attendees enjoyed a diverse program featuring 32 student presentations and insightful talks from three esteemed guest lecturers in sociolinguistics, forensic linguistics, and the world of language education. We also had daily panels, allowing attendees to learn about various career paths and postgraduate opportunities, offering invaluable insights for participants keen on charting their linguistic journey. But ULAB 2023 was not just about academics - it was about forging connections and experiencing the vibrant culture of Manchester. Each day offered social opportunities for attendees to mingle, creating a sense of community we have all come to love and hold dear to ULAB, and allowing exploration of our great city!

A particular personal achievement for me was being able to facilitate a truly great collaboration between the University of Manchester (official host institution) and Manchester Metropolitan University. With the invaluable support of Prof. Rob Drummond, whose work in sociolinguistics through his 'Manchester Voices' project can be seen displayed throughout public institutions in the city, helping to promote ULAB to his most motivated students, we ended up with a fabulous, dynamic, multi-institutional community. It strengthened relationships between linguistic departments of our universities and also the sense of community in general in our city. I would like to thank the Local committee for their insanely hard work ensuring the smooth running of the conference, and to staff at UoM and MMU for generous funding and support throughout the experience. I am very excited to see what Sussex achieves with their ULAB 2024 conference.

Introduction to the Proceedings

James Berriman, University of Portsmouth

It is an honour to present the Proceedings of the ULAB 2023 conference. As ULAB's 2024-25 archivist, I was given the task of finalising the 2023 Proceedings by former archivist and national chair Lydia Wiernik. This acted as a handover, allowing me to get to grips with the style guide and what is needed to format, edit, and publish a Proceedings document.

I was not personally involved with ULAB at this time, but what is clear to see from these Proceedings is that the organisation was thriving as it has done for its existence. Undergraduate students from across the length of Great Britain and beyond attended and presented at the conference at the University of Manchester in April 2023, showcasing some of the finest linguistics research on offer across the world. Students were able to choose between attending online and in person, meaning that all were given the opportunity to attend how they preferred.

Since 2019, ULAB has given conference presenters the opportunity to have their research published and it is so important to ensure that everyone that wishes to be published is included and gets their research out there. It is also vital to keep these sorts of resources to join the materials in our archive.

These Proceedings are comprised of five articles written by undergraduate students who presented their research at ULAB XIII at the University of Manchester. They are organised into two parts. Firstly, in Section A, full length papers that were assessed and graded as part of a university course, achieving a Class I or II.1, or international equivalent, can be found. Section B is made up of shorter squibs and write-ups. These papers are shorter assessed pieces and works conducted independently of university courses and were therefore ungraded. This method assures students access to publication regardless of length of paper or stage of study. All are of outstanding quality and span the many fields of study within Linguistics. In the Section B, we have ULAB history with a critical review being published for the first time in a conference Proceedings.

I am hugely grateful to many for their contributions to these Proceedings. Firstly, thanks to my predecessor Eden Elia Müller for the work in collecting the papers and beginning the editing process, as well as the fantastic cover design. My thanks also go to Lydia Wiernik for setting me this task and helping me with all the challenges I faced along the way. Thanks to the team of copyeditors who worked on these Proceedings: Nicholas Daines, Alice Duncan, Alice Eddyshaw, and Ilias Kolokousis. Finally, a huge thank you to the authors for their patience while they waited for their papers to be published.

Please enjoy the 2023 Proceedings of ULAB.

SECTION A

ASSESSED CONTRIBUTIONS

The following contributions have been assessed by academic institutions and have been awarded a Class I, Class II.1, or international equivalent.

Word Reading Processing by Bilingual Speakers: Brazilian Portuguese and English with and without ADHD

Débora Morais Barbosa da Silva
Federal University of Paraiba – Brazil

Abstract. The research explores the relationship between bilingualism and ADHD, based on the way bilingual people process word reading. Faced with the many benefits of bilingualism for the individual, we set out to investigate what would happen if we combined the two phenomena. It is a fact that the bilingual person has improved executive control. Therefore, would it be beneficial for the subjects with ADHD if they were bilingual? The hypothesis for our study is that bilingual participants with ADHD would perform equivalently to bilingual participants without ADHD. We carried out a self-monitored reading task of isolated words, to measure the reading time of the words and the index of correct answers. The material consists of a set of 48 word/non-word/pseudoword pairs. The dependent variables are reading time for the first word, reading time for the second word, response time for the lexical decision of the word/pseudoword/non-word and index of correct answers. The independent variables are frequency and length of words and group of bilinguals (with and without ADHD). The results indicated that the bilingual participants with ADHD read the first and second word in less time than the group without ADHD and had an equivalent correct index, presenting evidence that points to an improved executive control for the bilingual subjects.

Keywords: Language Processing; Bilingualism; ADHD; Executive control

1 Introduction

One of the central interests of Experimental Psycholinguistics is to understand how people process verbal language, and what are the mental processes involved in this processing in real time (Leitão, 2005). Within this, there are several specific interests, ranging from investigations at the phonological level to the semantic-pragmatic levels, both in the native language and in the second language (L2). Among the objects of study of Experimental Psycholinguistics, there is an interest in studying how the linguistic processing of words occurs.

In contrast to bilingualism, Attention Deficit Hyperactivity Disorder (ADHD) is associated with poor executive control (Barkley, 2020; DiGirolamo et al., 2001; Sonuga-Barke et al., 2010) and its main symptoms are: inattention, impulsivity, and hyperactivity (American Psychiatric Association, 2014). Onset of these symptoms typically occurs before age 12 years and has been present for at least six months, and in at least two settings. Children with ADHD have other difficulties in language development and academic functioning, such as, for example, an impoverished linguistic functioning, such as difficulty in retelling stories, producing long sentences, interaction skills in each context (Bellani et al., 2011).

ADHD is considered a public health problem, due to impairments in the cognitive, social, family, academic, and occupational spheres of people with this disorder, generating costs to society (Diller et al., 1996; Nevin, 2003). One of these impairments is in working memory, which can influence the speed of word recognition, due to the size of the sequence of letters or sounds (Beyersmann et al., 2011). The linguistic processing of reading demands the identification of words through the recognition of letter sounds and/or direct lexical access, converting graphic signs into phonological representations. The grapheme/phoneme conversion is initially carried out slowly, during literacy, characterized by a

syllabic reading rhythm that occurs by the phonological route (Cavalheiro et al., 2010). The reading process starts with visual processing, followed by letter decoding, a task that requires attention and concentration. For subjects with ADHD, this path may fail, as they have difficulty maintaining concentration, also impairing textual comprehension, making it necessary to repeat the reading.

In view of the studies showing the benefits of bilingualism for the individual, and the differences in processing for this group, we propose to investigate what happens if we combine the two phenomena. Would the possibility of an improved executive control for bilinguals then be beneficial for the subject with ADHD since they have a deficit in executive control? Can bilingualism act upon favouring the processing of reading isolated words compared to monolingual individuals with ADHD?

The main objective of this research was to investigate the reading time of words in bilingual L1-Brazilian-Portuguese and L2-English speakers with and without ADHD. The specific objectives aim to analyse the linguistic factors that influence the lexical processing (lexical decision - word vs. non-word and high vs. low frequency of a word) in bilinguals, to correlate the improvement of executive control in bilinguals with the attentional factor in ADHD in a self-monitored reading task with lexical decision, and to investigate whether the lexical linguistic frequency influences the reading time and the task among the tested groups. To do so, we conducted a psycholinguistic experiment comparing the performance of linguistic processing between bilinguals with and without ADHD.

This research is within the scope of Experimental Psycholinguistics regarding the linguistic processing of bilingual subjects with and without ADHD. Through observation through the experimental task of self-monitored reading of words with lexical decision, we will investigate the reading times of the words and the response.

2 Lexical Access

A variety of models have been proposed to explain how the bilingual mental lexicon is organised. Schwartz & Kroll (2006) observed that there is a hierarchical structure of words and concepts. The Word Association model says that ‘L2 words access meaning indirectly through the L1’ (op. cit.), i.e., the L2 words are represented in the bilingual's mind as the equivalent L1 translation. The Concept Mediation model proposes that ‘L2 words have direct access to their respective meanings’ (op. cit.).

The revised hierarchical model (RHM), proposed by Kroll & Stewart (1994), integrates the two previous models and is able to take into account changes in connections between words and concepts as L2 acquisition progresses. RHM assumes, at the lexical level, that ‘the L2 to L1 connection is stronger than the L1 to L2 connection’ (Schwartz & Kroll, 2006, p. 971) due to the stage at which learners use L1 translations to retrieve the meaning of L2 words. Therefore, L1 connections to concepts are stronger than L2 connections, but the model recognizes that, as language proficiency increases, L2 connections to concepts start to become like L1 connections. As a concept mediation model, RHM accounts for the possibility of skilled bilinguals accessing concepts directly through L2 words. Evidence for RHM comes from translation experiments and semantic categorization tasks (Dufour & Kroll, 1995). It is the first model to explain changes in mental representation during second language acquisition.

Dijkstra & Van Heuven (2002), present a connectionist model called BIA+, (Model of Bilingual Plus Interactive Activation) which aims to explain and simulate, through computational modelling, how the asymmetric patterns observed in bilingual processing occur. In this model, the lexicons of the bilingual's languages are incorporated into a single component, thus as he argues that there is no separation between languages during processing, that is, non-selective lexical access. As its name suggests, BIA+ is an expanded version of an existing model, the BIA (Van Heuven et al., 1998) which, in turn, is based on the monolingual Interactive Activation model of McClelland and Rumelhart (1981), whose distinction

between them lie in the simplification of the model, as well as the inclusion of elements not previously considered by the BIA. Basically, the model did not take into account external factors not linguistic elements, such as context, types of tasks and other aspects, as well as phonological and semantic elements. Due to the various limitations presented by the model, especially in relation to the description and functionalities of the linguistic nodes responsible for identifying the language corresponding to a given stimulus, and the lack of morphological and semantic representations, the need to modify it arose. Unfortunately, BIA and BIA+ only simulate spelling recognition of 4-letter (or 5-letter) words; the lexical-semantic representations are not implemented.

The most recent model is the Multilink (Dijkstra et al., 2019), a computational model for bilingual word recognition and word translation. This model considers the findings obtained with the BIA/BIA+ and RHM models and integrates them into the Multilink model. One of Multilink's objectives is to go beyond the spelling recognition of 4-letter (or 5-letter) words simulated by BIA and BIA+, simulating now the recognition of 3- to 8-letter words, including cognates of different lengths. While the BIA/BIA+ models consider word retrieval during language comprehension, RHM addresses language production issues, in particular word translation (Dijkstra & Rekké, 2010, pp. 403–407). According to the RHM, in the early stages of L2 acquisition, the meaning of words presented in the L1 is retrieved directly but retrieving that of the L2 words requires mediation through the L1 translation equivalent. This indirect mediation proceeds through word association links between representations of L2 and L1 word forms.

A word in one language is presented to a bilingual in a printed or spoken form who then produces its translation into another language, usually in spoken form. This 'translation production' task covers both reverse translation (L2 to L1) and direct translation (L1 to L2). In this task, participants must actively generate a response based on their knowledge of the foreign language (De Groot et al., 2000). On the other hand, in the translation recognition task, participants decide as quickly as possible whether the target is a correct prime translation or not. The Multilink model distinction is in activating their semantic representations in a way associated with the orthographic representations activated in the input (Dijkstra et al., 2019) and the semantic representations activate the non-selectively linked phonological representations of the language. As multiple lexico-phonological representations of the input and output language are co-activated, there must be a task-dependent decision process that ensures that the correct translation in the correct language is produced.

A computational model like Multilink helps us understand the mechanisms underlying the retrieval of bilingual words both qualitatively and quantitatively. Multilink sheds light on how various underlying mechanisms interact and function. For example, particular assumptions on the input or output side can be added or omitted to see if they are essential to generate specific Reaction Time patterns (e.g., asymmetry of translation effects). Thus, a computational model such as Multilink inspires the modeler to generate new hypotheses for empirical testing, as it allows simulations of complex interactions between many variables (Dijkstra et al., 2019).

Brysbaert & Duyck (2010, p. 364) presented evidence of 'L1 and L2 words acting very much as if they were words of the same language, interacting with each other as part of the word identification process', indicating little evidence for separate lexicons. In Interactive Activation (IA) models, languages can only be seen as separate to the extent that there is a difference in the size of lateral inhibition (between orthographic or lexical phonological nodes) within and between languages. In Multilink, it is assumed that there are no inhibitory side effects between words, nor between languages, as a kind of null hypothesis.

In Multilink model, in which connections between L2 words and their meanings are stronger, word forms are characterized by a frequency-dependent resting (waiting) level activation. As bilinguals with different L2 proficiency use L2 words more or less frequently, this implies differences in resting level activation for L2 words. Multilink was developed with the aim of providing an implemented general description of word form and meaning retrieval during word recognition and production (Dijkstra &

Rekké, 2010). Future developments of the model aim to explore, for example, whether the intensity of relationships between meaning and resulting sonority may depend on language proficiency.

Psycholinguistic models of the bilingual mental lexicon focus on which level of representation – orthographic/phonological, lexical or conceptual level – bilingual languages are interconnected (Isel et al., 2010). Understanding this phenomenon guides the investigations conducted with multilingual individuals, delimiting the starting and ending point of interference of the second language on the first, as well as assessing the real extent of this relationship.

3 Bilingualism

The phenomenon of bilingualism is the subject of much debate and discussion, not only regarding its definition, but also its relevance. As for its definition, it pervades the questions of how to know that the individual is bilingual, and how to determine their level of proficiency. Schwartz & Kroll (2006) define as bilingual individuals who actively use two languages with some degree of proficiency and explain that they rarely tend to be 'equally proficient or balanced in the use of two languages, making one of the languages the most dominant in the language' (p. 968). Researchers such as Grosjean (2012) define bilinguals as 'those who use two or more languages (or dialects) in their everyday lives' (p. 4). Furthermore, bilinguals do not form a homogeneous group, they vary in age and form of acquisition, level of proficiency and how much they use their languages.

Bilinguals can be dominant in one of their languages or balanced, but defining the notion of language mastery is difficult. The criteria for this definition are based on fluency, use, reading and writing skills, with fluency as a highlight. In the literature, it is possible to find subjective fluency, when participants self-report their fluency in languages in a background questionnaire; and objective fluency, when the researchers assess the fluency of the participants through assessment instruments, which are proficiency tests or evaluative tests carried out by the researchers themselves. Furthermore, bilinguals may not develop full and equal fluency in all language skills (speaking, listening, reading and writing) and their linguistic repertoire may change over time depending on the degree of use of one or more knowledge languages. Experimental research has suggested that both languages of a bilingual are activated jointly, even when the context does not require activation of both (Grosjean, 2012).

In this fertile field for investigative research on the possible benefits of bilingualism, studies such as Mårtensson et al. (2012) suggested that 'the learning of foreign languages in adults is accompanied by increases in the volume of grey matter in brain regions related to language' (p.244). Mechelli et al. (2004) found 'an increase in grey matter density in the left inferior parietal cortex of bilinguals relative to monolinguals that is more pronounced in early bilinguals than in late bilinguals' (p.757).

3.1 Bilingualism and Interfaces

3.1.1 Bilingualism and Executive Control

Executive functions are cognitive processes that control behaviour in the service of goal achievement (Diamond, 2013). These cognitive skills are essential for planning behaviour, ignoring irrelevant information, attending to stimuli and information of interest, and creativity. Executive function skills change throughout life because of cognitive maturation and after age-related decline (Dempster, 1992). They are also malleable and respond to short- and long-term experiences (Diamond & Lee, 2011), which leads us to infer that bilingualism affects executive functioning in some way, given that constant monitoring, inhibition, selection, and planning are essential components of everyday bilingual language

use. What can be expected is that bilingual language processing will rely more on general domain executive functions than in the case of monolinguals, and transfer effects will give bilingual advantages in executive functions, according to results from tests such as Stroop, Simon, and Flanker Task.

An experiment performed by Bialystok et al. (2017) with bilingual children with ADHD presented evidence for a more laborious linguistic processing in bilingualism, which means a reduced vocabulary in each language and slower word retrieval. However, aspects of cognitive processing, in particular executive control, have been improved. ADHD was associated with a weakened executive control system, with symptoms of inattention, impulsivity and hyperactivity. Language, Flanker, and Stop signal proficiency tests were taken and an interesting result has been presented regarding language proficiency; the one-way ANOVAs comparing the two bilingual groups showed that bilinguals with ADHD, rated themselves more proficient than non-ADHD bilinguals in English proficiency. There was no self-reported evidence of language proficiency disadvantages for bilingual individuals with ADHD, indicating that in the college student sample, ADHD may be associated with high verbal functioning.

A few years later, Chung-Fat-Yim et al. (2019) present a study pointing to the impact of bilingualism on executive control, now with bilingual adolescents performing better on tasks than monolinguals. As we saw in Section 2.1, research has shown that even when only one language is needed, bilinguals activate the lexicons of both languages in parallel (Kroll et al., 2012). Due to this joint activation, the bilingual brain must manage language selection by focusing its attention on the language in question, and in return, ignore the interference of the other language, which presumably occurs through the recruitment of general attention mechanisms (Bialystok, 2015). The unique experience of bilinguals in managing attention to two conjointly activated languages is training in selective attention, a crucial element of executive function. Bilingual adolescents showed higher levels of executive functioning than their monolingual peers on the standard flanker task, a result also found with children. Bilingualism has been linked to improved attention skills.

3.1.2 Bilingualism and Brain Plasticity

Throughout this research, we have presented evidence that bilingualism acts on neural networks, and it is widely accepted that this phenomenon changes the brain (Pliatsikas et al., 2014b). These changes in brain structure are indicators of experience-dependent plasticity and impact brain function and thus executive functions. Bilingualism affects brain regions that serve cognitive control, including the left inferior frontal gyrus, anterior cingulate cortex, inferior parietal lobe, and basal ganglia, particularly the putamen and left caudate nucleus (Abutalebi & Green, 2016). Bilinguals have greater grey matter density than monolinguals in several brain structures, including the left inferior parietal lobe, which is modulated by age of acquisition and proficiency (Mechelli et al., 2004), as well as the caudate nucleus (Zou et al., 2012) and the cerebellum (Pliatsikas et al., 2014b). Bilinguals also have greater white matter integrity than monolinguals (Luk et al., 2011).

Rose et al. (2013) studied 15 bilingual children for one year and found that inferior parietal lobe grey matter density increased and was related to language ability and cognitive control. Gold et al. (2013) observed that older adult bilinguals had lower blood oxygen level-dependent response (indicative of processing with less effort) than monolinguals in several frontal regions and exhibited superior task-switching skills. Zou et. al. (2012) found that grey matter volume of the anterior cingulate cortex (an area implicated in executive control) in bilinguals was positively correlated with functional activity and negatively correlated with a behavioural conflict effect. A few years later, Abutalebi et al. (2015) in a study on the neuroprotective effects of bilingualism with older Chinese adults, reported increased grey matter in the anterior cingulate cortex, while monolinguals showed decreased grey matter in the dorsolateral prefrontal cortex (which plays a crucial role in executive functions), and these brain

differences correlated with the advantage of bilinguals over monolinguals on the Flanker task. These findings suggest that bilinguals benefit from more efficient executive function processes and that this can be observed in the anatomical correlates of the processes in question.

Cognitively stimulating activities, both long term and short term, lead to cognitive benefits, brain changes and improved cognitive outcomes of aging. Bilingualism is one such cognitive stimulus, and possibly involves a significantly larger brain network than, say, crosswords or Sudoku puzzles or learning to play an instrument (Antoniou, 2019). This evidence supports the view that bilinguals have an advantage in offsetting age-related decline.

4 Attention Deficit Hyperactivity Disorder (ADHD)

According to the Diagnostic and Statistical Manual of Mental Disorders (DSM-5, 2014), ADHD is:

> '...a neurodevelopmental disorder defined by impairing levels of inattention, disorganization, and/or hyperactivity-impulsivity. Inattention and disorganization involve an inability to stay on task, an appearance of not listening, and loss of materials at levels inconsistent with age or developmental level. Hyperactivity-impulsivity implies excessive activity, restlessness, inability to remain seated, intrusiveness in activities of others, and inability to wait - symptoms that are excessive for age or developmental level. In childhood, ADHD often overlaps with disorders commonly considered "externalizing," such as oppositional defiant disorder and conduct disorder. ADHD often persists into adulthood, resulting in impaired social, academic, and professional functioning.'

ADHD is the most common neurodevelopmental disorder of childhood (Fontana et al., 2007). Some studies reveal that ADHD is a neuropsychiatric condition that affects 3-5% of children and that often persists into adulthood, considered one of the most common neuropsychiatric disorders. The electronic magazine 'Consciência no dia a dia' (2009), explains that attention deficit happens because neurotransmitters such as dopamine (main component of our cerebral reward system, a system activated every time we do something that gives pleasure and signals the brain that the experience is worth repeating) do not work as well in individuals with ADHD.

For Olivier (2010), its diagnosis can also be defined as a multifactorial disorder associated with environmental and genetic factors. Variations in brain size and morphology, where they are present very early, with abnormalities in the frontal/cerebellum circuit, mainly in the right hemisphere, thought to be responsible for most disorders of motor coordination and a subnormal sensorimotor program, in addition to other possible causes, but not widely publicised.

Inattention is a difficulty in paying attention to details or making careless mistakes in certain activities such as: difficulty maintaining attention in tasks and recreational activities, seeming not to listen when spoken to, not following instructions and not finishing tasks, difficulty organising tasks and activities, avoid engaging in tasks that require constant mental effort, losing things necessary for tasks or activities, and being easily distracted by stimuli unrelated to the task and showing forgetfulness in daily activities (DSM-5, 2014, p. 59) .

Hyperactivity refers to excessive motor activity - frequent fidgeting of hands and feet, fidgeting a lot in a chair, abandoning the chair in situations in which it is expected to remain seated, talking too much, difficulty engaging silently in activities, running in inappropriate places, giving hasty answers before questions have been completed, have difficulty waiting their turn, and frequently interrupting or meddling in other people's business (DSM-5, 2014, p.60).

Impulsivity refers to hasty actions that occur at the moment without premeditation and with a high potential for harm to the person, some common situations: crossing a street without looking, may manifest with social intrusiveness, making important decisions without consideration of long-term consequences, lapses in dispersion may generate difficulty in temporal organisation, interpersonal relationships, leading the person to generate too much effort to carry out everyday tasks and resulting in less durability and performance, difficulty playing sports, excessive forgetfulness of everyday activities such as turning off the gas, among others, difficulty maintaining a job, social group, or relationship for a long period of time, verbal impulsiveness and/or behaviours of trying to defend any action, showing aggressive behaviour, uncontrolled eating, drug use, excessive spending, varied compulsions such as for games, uncontrollable chatter, among others; they will be able to participate in several projects at the same time, affirming the need to live intensely (DSM-5, 2014, p.60).

In view of the symptoms observed for at least six months, individuals with ADHD can be classified into three subtypes: mixed or combined, predominantly hyperactive and predominantly inattentive (Fontana et al., 2007). The diagnosis of ADHD is fundamentally clinical, based on well-defined operational criteria, derived from classification systems such as the DSM-4 or the ICD-10. The DSM-4 proposes the need for at least six symptoms of inattention and/or six symptoms of hyperactivity/impulsivity for the diagnosis of ADHD. The DSM-4 and ICD-10 include a criterion of age of onset of symptoms prejudicing (before seven years old) for the diagnosis of the disorder.

4.1 ADHD and Executive Functions

The concept of executive function begins with the neuropsychological study of patients with deficiencies in this area, as well as the formulation of working memory. Baddeley (1996) organized the working memory model as a tripartite storage system consisting of a central executive and two slave systems: a phonological loop, related to the representation and recitation of verbal material, and a visuospatial buffer system, the imagery equivalent of memory. The central executive is the main component of the theory, and its functions are reasoning, decision-making, strategy planning and behaviour control through the integration of information from subordinate systems. Its operation involves learning and applying contingent rules, abstract reasoning, and maintaining attention and concentration, and its functioning is closely related to the prefrontal areas.

Van Lambalgen et al. (2008) present causal relationships between the etiology of ADHD and executive function, showing deficiencies in planning. The sample consisted of a group of 26 children with ADHD and 34 individuals without ADHD, divided according to the age group of seven to nine years. All children with ADHD categorized as 'combined type', i.e. with attention and hyperactivity problems. An important factor is the presence of comorbidity in 22 of 26 children with ADHD. For testing executive functions, the study was based on the task performed by Shue et al. (1992) who showed that children with ADHD score significantly worse than neurotypical individuals in a set of tasks involving synthesis and execution of a plan to reach a certain goal. The paradigmatic case is the 'Go/No-Go ' task, whose basic structure is as follows: in one type of stimulus the subject has to press a key (this is the 'go' stimulus), in another type of stimulus the subject must not do anything (the 'no go' stimulus). Stimuli are presented in random order. In Shue et al. (1992), the 'go' stimulus is a paper showing an apple, and the 'no go' stimulus is a paper showing an ice cream; The goal is 'do (x) now', with x a variable to be instantiated. There are two rules, which can be formalised as 'if it's an apple, go (go)' and 'if it's an ice cream, go (don't go)'. This means that there are two possible unifications for the variable x in each goal, the go, and don't go actions, which reduce the goal to satisfying one of the preconditions, apple, or ice cream. The (only) successful unification then determines the action to be taken. Performance on this task in children with ADHD is significantly impaired, as many attempts to 'no go' lead to a 'go' response, and this suggests that the initial

calculation (unification and reduction of the goal 'do (x) now') is not executed correctly. If the 'go' answer is conceived as an unconditional answer, that is, if the goal is simplified to 'do (go)', it does not require any calculation. Thus, we see the performance of children with ADHD on these tasks as evidence of the difficulty of keeping a complex goal active in working memory and the tendency to simplify that goal as a result.

In the production of utterances, skills such as semantics and pragmatics in subjects with ADHD are affected. In a 'story telling' task, whose objective is to investigate which linguistic devices children use to narrate the story, subjects with ADHD present deviant narration patterns due to executive difficulties. These difficulties are corroborated by the study by Purvis & Tannock (1997) that showed in the investigation of this phenomenon with the reading of a popular tale 'The father, his son and their donkey' in which children would have to repeat the story, that there was a tendency to retell the events of the story out of sequence.

ADHD, within what was researched, is a neurological disorder and it is estimated that the causes are associated with the non-functioning of neurotransmitters such as dopamine, genetic factors, and the history of childhood. Difficulty in concentration and attention, restlessness in places like the classroom, lectures, etc., too talkative, impatient and distracted, are some of the characteristics of this group. The consequence of these symptoms is perceived in the daily life of these subjects and a lot of intervention needs to be done, both pharmacologically and through therapies.

5 Experimental Design

Although ADHD is one of the most frequent disorders in childhood and consequently one of the most studied in the child population, there are still few studies that aim to relate the cognitive function of working memory with the areas of learning difficulty in attention deficit disorder/hyperactivity in both children and adults, which are sometimes diagnosed late.

We conducted a self-monitored reading to observe how words are processed in English by bilingual subjects (Portuguese – English) with and without ADHD, at intermediate and advanced levels of proficiency. According to Mitchell (2004), the self-monitored reading technique is one of the techniques, along with eye tracking and the method that makes use of event-related evoked potentials (ERPs) through electroencephalograms (EEGs), that are timeless. In the self-paced reading technique, a subject is presented with text segmented into words, strings of words or phrases that are displayed one at a time on a computer screen. The individual himself is responsible for pressing a key so that the text or word begins to be shown, and after reading it press the key again to show the second part of the text or word and so on until the end of the task. The experiment was performed using the PennController for Internet-Based Experiments software ('PennController' or 'PCIbex' for short) which provides the tools to build and run experiments online, from familiar paradigms such as self-paced reading to completely customized paradigms.

The hypotheses we assume are that: Bilinguals with ADHD will perform similarly to bilinguals without ADHD in the self-monitored reading task with lexical decision (word versus non-word) modulated by lexical frequency - high versus low). We justify these hypotheses based on recent studies regarding the improvement on executive control of bilingual people and the better performance of ADHD bilingual people in their Language and Flanker test (Bialystok et al., 2017; Chung-Fat-Yim et al., 2019). Bilinguals will have a cognitive advantage and better performance compared to monolingual controls without ADHD.

14 volunteer subjects[1] participated in this experiment, Brazilian English speakers, seven subjects with ADHD, and seven without ADHD (control group). Within the group with ADHD, we have four subjects at the advanced level and three subjects at the intermediate level, as well as in the group without ADHD. The age between 25 and 40 years old, among the forwards we have three residents from João Pessoa, two from São Paulo, two from England, and one from Portugal. In the group of intermediaries, we have two from São Paulo, three from João Pessoa, and one from Canada. Of these participants, six are women and eight are men. There was the intention to control the medicated/non-medicated factor of each volunteer with ADHD, and so we conducted a questionnaire on Google Forms so that they could answer these questions: 1. At what stage was it diagnosed? (childhood, adolescence, adult), 2. Do you use medication? If yes, how long ago? 3. What strategies do you consciously use to deal with the disorder?

We also recorded any behavioural changes observed by them on the day the experiment was carried out, such as anxiety, fatigue, excessive lack of concentration, etc., due to external circumstances, and we instructed them to carry out the experimental task at another time, to be able to explain any time and correct answers given by participants with ADHD. Only one participant communicated that they were not feeling very well that day, possibly having an anxiety attack, and carried out the experiment two days later.

When collecting the responses from the form, we found that of the seven participants with ADHD, six were diagnosed in adulthood and had been using medication for about two to five years. Among the main responses to which strategies they use to deal with the disorder were: use of reminders, support network, physical exercise, routine, avoiding exposure to excessive stimuli, regulated sleep, controlled caffeine intake, meditation, task management, psychotherapy, paying close attention to everything you are doing, and coping cards.

5.1 Procedures

Initially, participants were informed that the research was within the scope of Experimental Psycholinguistics, and that they could withdraw from participating at any time, without prejudice. They were also informed about how many steps would be performed by them, and the approximate total time of these procedures, which was 20 to 25 minutes. In order to evaluate their proficiency level, participants answered the proficiency test, the Vocabulary Levels Test (Nation, 1990), through the *ClassMarker* platform.

The experimental task is based on experiment three of the thesis of Albuquerque (2008). The online experiment was carried out using the isolated word self-monitored reading technique, which, in this case, consisted of carefully reading two words comparing their reading times and the rate of correct answers between the groups in question. Independent variables are word frequency and length, and a bilingual group (with or without ADHD). The dependent variables are, therefore, the reading times of the words, the lexical decision time for the second word/pseudoword/non-word and the index of correct answers. The hypothesis for this experiment is that the group with ADHD will present reading times equivalent to those of the group without ADHD.

The material consisted of a set of 192 words separated by conditions: 48 high-frequency small word/pseudoword/nonword pairs, 48 high-frequency big word/pseudoword/nonword pairs, 48 pairs of low-frequency small words/ pseudowords /nonwords, and 48 low-frequency big word/pseudoword/nonword pairs. The words were taken from the Academic Word List (AWL). Since the separation of syllables in the English language has its specifications, we chose to control the size of the

[1] Another 16 participants will still be tested to complete the sample of 30 participants.

words by the number of letters, so words with four to six letters would be considered 'small' and those with seven to nine letters, 'big'.

5.2 Results

The dependent variables of the experiment were the time to read the first word (RT1), the time to read the second word (RT2), the time to answer the question at the end of the task (RT3). If the words were the same, the participants would press the F key and if they were not the same, they would press the J key. This procedure was reversed for the participants in the second round of the experiment, if the words were the same, they would press the J key and if they were not, the letter F. The last dependent variable was the hit rate.

We selected 14 participants for the first round of the experiment, separated by group and proficiency. This sample contained three participants from the control group and three participants from the intermediate-level ADHD group; and four participants in the control group and four participants in the advanced ADHD group. First, we carried out the descriptive statistics in the Jamovi software of all the reading times RT1, RT2, and RT3 and the index of correct answers in relation to all independent variables. We remove outliers using three standard deviations up and down.

Then, we performed an analysis of variance test (ANOVA) in relation to the experiment items and the variable reading time of the first word. For the dependent variable RT1, the analysis showed a main effect between groups, with and without ADHD with F values (1.298) = 116.095 and $p < .001$ and for the length of words with F values (1.298) = 6.984 and $p = 0.008$. There was no effect on proficiency ($p= 0.402$) or word frequency ($p=0.485$). There was interaction factor between group ✻ proficiency ✻ size with values F (1.298) = 6.974 and $p = 0.008$ and group interaction ✻ proficiency with F values (1.298) = 10.327 and $p= 0.001$.

Post Hoc Tests of *t*-values show us exactly where the interactions identified by the ANOVA between group ✻ proficiency occurred. The values show that the advanced ADHD group compared to the advanced control presented a shorter reading time with 329 milliseconds of average difference, with $t = 5.72$ and $p<0.001$. The intermediate ADHD group compared to the advanced control showed a mean difference of 505 milliseconds in favour of the ADHD group with $t = 8.18$ and $p<0.001$. The intermediate control and advanced ADHD group differed by 432 milliseconds with $t = 7.06$ and $p<0.001$. The intermediate control and intermediate ADHD group was where we found the greatest statistical difference with $t = 9.32$, $p<0.001$ and mean difference of 608 milliseconds. We did not find statistically significant differences between the groups, ADHD and control and their levels of proficiency, that is, advanced control group and intermediate control and advanced ADHD and intermediate ADHD group. The results showed shorter reading times in the RT1 variable for the bilingual subjects with ADHD compared to the control group, which we expected according to our hypothesis, equivalent reading times between the tested groups.

Then descriptive statistics, ANOVA and *t*-test for the reading time of the second word, RT2, were performed. For the dependent variable RT2, the analysis showed a main effect between the control and ADHD groups with F values (1.277) =33.7006, mean difference of 142, $t=5.53$ and $p< .001$ and for the proficiency level F (1.277) = 42.6220, mean difference of 174, $t=6.81$ and $p< .001$. There was interaction effect between group ✻ proficiency with F (1.277) = 18.5447 and $p<0.001$, group ✻ proficiency ✻ size with values of F (1.277) =5.5631 and $p= 0.018$, group ✻ proficiency ✻ frequency with value of F (1.277) = 4.3170 e $p=0.038$, group ✻ size ✻ frequency with F (1.277) = 12.1782 and $p<0.001$ and finally group ✻ proficiency ✻ size ✻ frequency with F(1.277) = 5.1854 and $p=0.023$.

We observed a greater effect between groups and proficiencies regarding large words for the intermediate level, the control-intermediate- nounbig condition and the ADHD-intermediate- nounbig group with a mean difference of 290.46, t=5.499 and p<0.001, a greater distancing of the average reading time between these groups. The same for small words, the control-intermediate- nounsmall condition and the ADHD-intermediate- nounsmall group presented a mean difference of 219.35, t=4.174 and p<0.001. The condition control and advanced ADHD nounsmall showed no statistically significant difference, as well as control and advanced ADHD nounbig.

For the third dependent variable, RT3, response time to the question asked after reading the words *prime* and *target* was also analysed, and we obtained the following results. We performed an analysis of variance test (ANOVA) for the dependent variable RT3, the analysis showed a main effect between groups, ADHD and control, and the level of proficiency. With values for the group effect of F (1,328) = 40,295 and p< .001, and for the level of proficiency we have values of F (1,328) = 147,534 and p< .001. ANOVA showed interaction between group ✻ proficiency with values and F (1,328) = 62,558 and p< .001.

The last dependent variable analysed was the rate of correct answers in relation to the group, proficiency, size and frequency, with the following results. For the index of correct answers between the groups, we found a statistically significant difference only between the proficiency levels, advanced and intermediate, with values F (1.328) = 2.9745 and p= 0.085 and interaction between group ✻ frequency ✻ size with values of F (1.328) = 5.3005 and p = 0.021. We observed that the intermediate-level ADHD group showed better performance in reading large words in relation to the advanced and the control group, both advanced and intermediate. This data is important for us to infer that, as we have a more favourable reading average for the ADHD group, their hit rates balanced with that of the control group, it can demonstrate that they were performing the task with attention, using their personal strategies.

5.3 Discussion

We proposed with our experiment to point out possible favours of bilingualism in the linguistic processing of typical subjects with ADHD, under the same experimental line. Aware of the impairment of ADHD with the executive function interface presented in the theoretical foundation of this research, we tested bilingual subjects, speaking Brazilian Portuguese-English, with and without ADHD and found in our sample of 14 participants, seven bilinguals with ADHD and seven bilinguals without ADHD, that participants with ADHD performed better on the experimental task. Showing shorter reading times than the control group and a balanced index of correct answers with participants without ADHD. We have lower reading time values towards the intermediate level, which we believe may be related to our sample size so far, of only 14 participants. Given the linguistic impairment that people with ADHD show in several studies, due to lack of attention, agitation, difficulty completing a task, etc., the findings found in this research seem to diverge from our initial hypothesis, that bilingual participants with ADHD would present equivalent performance with bilinguals without ADHD, but we are still moving forward with the collection of data from the other participants.

The results seem to indicate bilingualism as a compensatory tool, as we saw in the study by Bialystok et al. (2017) with bilingual children with ADHD who, despite the reduced vocabulary and weakened executive control due to the disorder, the linguistic proficiency in English of these participants was higher than bilinguals without ADHD. Evidence supported by Chung-Fat-Yim et al. (2019) pointed out the impact of bilingualism on executive control, now with bilingual adolescents, who performed tasks better than monolinguals. Bilingual adolescents showed higher levels of executive functioning than their

monolingual peers on the standard flanker task, a result also found with children. Bilingualism has been linked to improved attention skills.

Another possible explanation that we based on why participants with ADHD performed the experimental task in less time than participants without ADHD is found in Messina et al. (2009), who performed five subtests with participants with ADHD and in the Visual Working Memory subtest, observed that the ADHD group performed better than the group without ADHD and that it may be due to the effort to remain attentive during the execution of the task, making use of from other areas that arise to help them in the attentional process. In view of the responses, we obtained from the participants regarding the strategies they use daily to better deal with the symptoms of the disorder.

An interesting fact regarding the rate of correct answers of our participants is that the answers were equivalent between both groups, the same result found in Albuquerque (2008), where the subjects with ADHD did not differ from the subjects without ADHD regarding the score of the answers; they just needed more time to perform the task, but they did not have a low rate of correct answers.

From the results obtained in this investigation so far, we have evidenced that the bilingualism factor may be acting favourably in the executive control of participants with ADHD. So that we can better contribute to its relevance, we will continue our research with the other participants and analysis of the variables and their respective statistical values. In the future of the investigation of the bilingualism factor, other issues could arise to be studied together, such as dyslexia and language disorders, with the aim of presenting more evidence and materials in this field of research.

6 Conclusion

The objective of this research was to compare the reading times of isolated words between the tested groups, bilinguals with and without ADHD. With our experiment, we proposed to point out possible benefits of bilingualism in the linguistic processing of individuals with ADHD, under the line of Experimental Psycholinguistics. We tested bilingual subjects, Brazilian Portuguese and English, with and without ADHD and found in our initial sample that participants with ADHD had a better response time in the experimental task. Presenting reading times shorter than the control group and correct answer rate balanced with participants without ADHD. Given the linguistic impairment that people with ADHD present in several studies, due to lack of attention, agitation, difficulty finishing a task, etc., the findings found in this research seem so far to diverge from our initial hypothesis, that bilingual participants with ADHD would present equivalent performance compared to bilinguals without ADHD, as in fact they had a better performance and conclude that there are no reading processing problems in the current sample of participants with ADHD. And we add evidence that the bilingualism factor may be acting in a compensatory way in the executive control of participants with ADHD.

We find limitations to our work regarding the model Multilink. The model is capable of explaining the issues of proficiency and lexical access, however we do not have a reference to explain some answers regarding the ADHD and this is a problem that we can continue to work on, as well as the investigation on bilingualism and language impairments.

7 References

Abutalebi, J. (2008). Neural aspects of second language representation and language control. *Acta psychologica*, 128(3), 466-478.

Abutalebi, J., Canini, M., Della Rosa, P. A., Green, D. W., & Weekes, B. S. (2015). The neuroprotective effects of bilingualism upon the inferior parietal lobule: a structural neuroimaging study in aging Chinese bilinguals. *Journal of Neurolinguistics*, 33, 3-13.

Abutalebi, J., & Green, D. W. (2016). Neuroimaging of language control in bilinguals: neural adaptation and reserve. *Bilingualism: Language and cognition*, 19(4), 689-698.

Albuquerque,G. (2008). *Language processing in attention deficit hyperactivity disorder* **[**Doctoral thesis]. Rio de Janeiro: UFRJ/Faculty of Letters.

American Psychiatric Association. (2014). *DSM-5: Manual diagnóstico e estatístico de transtornos mentais*. Artmed Editora.

Antoniou, M. (2019). The advantages of bilingualism debate. *Annual Review of Linguistics*, 5(1), 395-415.

Baddeley, A. (1996). Exploring the central executive. *The Quarterly Journal of Experimental Psychology Section A*, 49(1), 5-28.

Barkley, A. (2020). *Attention Deficit Hyperactivity Disorder Handbook* (3rd Ed.). London: The Guilford Press.

Bellani, M., Moretti, A., Perlini, C., & Brambilla, P. (2011). Language disturbances in ADHD. *Epidemiology and Psychiatric Sciences*, 20(4), 311-315.

Beyersmann, E., Castles, A., & Coltheart, M. (2011). Early morphological decomposition during visual word recognition: Evidence from masked transposed-letter priming. *Psychonomic Bulletin & Review*, 18, 937-942.

Bialystok, E., Luk, G., Peets, K. F., & Sujin, Y. A. N. G. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and cognition*, 13(4), 525-531.

Bialystok, E. (2015). Bilingualism and the development of executive function: The role of attention. *Child development perspectives*, 9(2), 117-121.

Bialystok, E., Hawrylewicz, K., Wiseheart, M., & Toplak, M. (2017). Interaction of bilingualism and attention-deficit/hyperactivity disorder in young adults. *Bilingualism: Language and Cognition*, 20(3), 588-601.

Brysbaert, M., & Duyck, W. (2010). Is it time to leave behind the Revised Hierarchical Model of bilingual language processing after fifteen years of service?. *Bilingualism: Language and cognition*, 13(3), 359-371.

Cavalheiro, L. G., Dos Santos, M. S., & Martinez, P. C. (2010). PC Influence of phonological awareness on reading acquisition. *CEFAC Magazine*, 12(6).

Chung-Fat-Yim, A., Himel, C., & Bialystok, E. (2019). The impact of bilingualism on executive function in adolescents. *International Journal of Bilingualism*, 23(6), 1278-1290.

De Groot, A. M., Delmaar, P., & Lupker, S. J. (2000). The processing of interlexical homographs in translation recognition and lexical decision: Support for non-selective access to bilingual memory. *The Quarterly Journal of Experimental Psychology: Section A*, 53(2), 397-428.

Dempster, F. N. (1992). The rise and fall of the inhibitory mechanism: Toward a unified theory of cognitive development and aging. *Developmental review*, 12(1), 45-75.

Diamond, A. (2013). Executive functions. *Annual review of psychology*, 64(1), 135-168.

Diamond, A., & Lee, K. (2011). Interventions shown to aid executive function development in children 4 to 12 years old. *Science*, 333(6045), 959-964.

DiGirolamo, G. J., Kramer, A. F., Barad, V., Cepeda, N. J., Weissman, D. H., Milham, M. P., ... & McAuley, E. (2001). General and task-specific frontal lobe recruitment in older adults during executive processes: a fMRI investigation of task-switching. *Neuroreport*, 12(9), 2065-2071.

Dijkstra, T. O. N., Wahl, A., Buytenhuijs, F., Van Halem, N., Al-Jibouri, Z., De Korte, M., & Rekké, S. (2019). Multilink: A computational model for bilingual word recognition and word translation. *Bilingualism: Language and Cognition*, 22(4), 657-679.

Dijkstra, T., & Rekké, S. (2010). Towards a localist-connectionist model of word translation. *The Mental Lexicon*, 5(3), 401-420.

Dijkstra, T., & Van Heuven, W. J. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and cognition*, 5(3), 175-197.

Diller, L. H., Tanner, J. L., & Weil, J. (1996). Etiology of ADHD: Nature or nurture?. *American Journal of Psychiatry*, 153(3), 451-a-452.

Dufour, R., & Kroll, J. F. (1995). Matching words to concepts in two languages: A test of the concept mediation model of bilingual representation. *Memory & cognition*, 23(2), 166-180.

Fontana, R. D. S., Vasconcelos, M. M. D., Góes, F. V. D., & Liberal, E. F. (2007). ADHD prevalence in four Brazilian public schools. *Arq. Neuro-Psiquiatr.*, 65(1), 134-137.

Grosjean, F. (1998). Studying bilinguals: Methodological and conceptual issues. *Bilingualism: Language and cognition*, 1(2), 131-149.

Isel, F., Baumgaertner, A., Thrän, J., Meisel, J. M., & Büchel, C. (2010). Neural circuitry of the bilingual mental lexicon: Effect of age of second language acquisition. *Brain and cognition*, 72(2), 169-180.

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of memory and language*, 33(2), 149-174.

Kroll, J. F., Van Hell, J. G., Tokowicz, N., & Green, D. W. (2010). The Revised Hierarchical Model: A critical review and assessment. *Bilingualism: Language and Cognition*, 13(3), 373-381.

Kroll, J. F. (2008). Juggling two languages in one mind. *Psychological Science Agenda, American Psychological Association*, 22(1).

Luk, G., Bialystok, E., Craik, F. I., & Grady, C. L. (2011). Lifelong bilingualism maintains white matter integrity in older adults. *Journal of Neuroscience*, 31(46), 16808-16813.

Mårtensson, J., Eriksson, J., Bodammer, N. C., Lindgren, M., Johansson, M., Nyberg, L., & Lövdén, M. (2012). Growth of language-related brain areas after foreign language learning. *NeuroImage*, 63(1), 240-244.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological review*, *88*(5), 375-407.

Mechelli, A., Crinion, J. T., Noppeney, U., O'Doherty, J., Ashburner, J., Frackowiak, R. S., & Price, C. J. (2004). Structural plasticity in the bilingual brain. *Nature*, 431(7010), 757-757.

Messina, L. D. F., & Tiedemann, K. B. (2009). Assessment of working memory in children with attention deficit hyperactivity disorder. *Psicologia USP*, 20, 209-228.

Mitchell, D. C. (2004). On-line methods in language processing: Introduction and historical review. In *The on-line study of sentence comprehension* (pp. 15-32). London: Psychology Press.

Nation, I. S. P. (1990). *Teaching & learning vocabulary*. Boston, MA: Heinle & Heinle Publishers.

Nevin, S. T. (2002). *Diagnostic efficiency of a battery of cognitive and activity measures for classification of childhood ADHD*. Virginia Consortium for Professional Psychology (Old Dominion University, College of William and Mary, Eastern Virginia Medical School, & Norfolk State University).

Pliatsikas, C., Johnstone, T., & Marinis, T. (2014a). fMRI evidence for the involvement of the procedural memory system in morphological processing of a second language. *PloS one*, 9(5), e97298.

Pliatsikas, C., Johnstone, T., & Marinis, T. (2014b). Grey matter volume in the cerebellum is related to the processing of grammatical rules in a second language: a structural voxel-based morphometry study. *The Cerebellum*, 13, 55-63.

Purvis, K. L., & Tannock, R. (1997). Language abilities in children with attention deficit hyperactivity disorder, reading disabilities, and normal controls. *Journal of abnormal child psychology*, 25, 133-144.

Schwartz, A. I., & Kroll, J. F. (2006). Bilingual lexical activation in sentence context. *Journal of memory and language*, 55(2), 197-212.

Shue, K. L., & Douglas, V. I. (1992). Attention deficit hyperactivity disorder and the frontal lobe syndrome. *Brain and cognition*, 20(1), 104-124.

Sonuga-Barke, E., Bitsakou, P., & Thompson, M. (2010). Beyond the dual pathway model: evidence for the dissociation of timing, inhibitory, and delay-related impairments in attention-deficit/hyperactivity disorder. *Journal of the American Academy of Child & Adolescent Psychiatry*, 49(4), 345-355.

Van Heuven, W. J., Dijkstra, T., & Grainger, J. (1998). Orthographic neighborhood effects in bilingual word recognition. *Journal of memory and language*, 39(3), 458-483.

van Lambalgen, M., van Kruistum, C., & Parigger, E. (2008). Discourse processing in attention-deficit hyperactivity disorder (ADHD). *Journal of logic, language and information*, 17, 467-487.

Zou, L., Ding, G., Abutalebi, J., Shu, H., & Peng, D. (2012). Structural plasticity of the left caudate in bimodal bilinguals. *Cortex*, 48(9), 1197-1206.

Morphosyntactic Analysis of Cypriot Turkish-English Codeswitching: Agglutination of Embedded Lexemes and Bilingual Periphrastic VP Creations

Arcin Celikesmer
University of York

Abstract. This study investigates agglutination and periphrasis in bilingual noun phrases (NPs) and verb phrases (VPs) when codeswitching between agglutinative Cypriot Turkish and synthetic English. The focus is on English elements embedded in Cypriot Turkish to assess adherence to the morphosyntactic rules of the Matrix Language (Myers-Scotton, 2006). Drawing on the 4M-Model (Myers-Scotton, 2006) and Boumans' (2007) analysis of bilingual VP patterns, the research examines the morphosyntactic behavior in Cypriot Turkish-English codeswitching, diverging from previous studies on Turkish-English pairings (Koban, 2013; Kemaloglu-Er, 2018). The study uses data from 90 minutes of semi-structured conversations with young Cypriot Turkish-English speakers, yielding 143 tokens of codeswitching, 49% of which were intra-sentential. Contrary to Poplack's (1980) claim that embedded words appear in bare forms, 57% of embeddings were agglutinated with up to two suffixes for nouns/adjectives and up to four for verbs. Predominant morphemes included Turkish dative, accusative, and pluralization suffixes, as well as possessive inflections. Bilingual VPs, comprising 18% of the tokens, often utilized periphrasis (61%) with light verbs such as "yap" (make), "et" (do), and "ol" (be). Embedded verbs also featured verb-making suffixes like "-le(mek)" (to do) and "-dı" (to be.PST). The findings largely support the MFL and 4M-Model, highlighting a correlation between intense language contact and periphrastic verb formations. Additionally, Cypriot Turkish shows similarities with Cypriot Greek rather than Turkic languages, with codeswitches behaving like assimilated lexical borrowings. This research underscores the role of codeswitching as an identity-building tool among young Cypriot Turkish-English bilinguals, reflecting their positive attitude towards bilingualism.

Keywords: Cypriot Turkish; bilingualism; codeswitching; morphosyntax; periphrasis; agglutination

---

1 Introduction

The morphological-syntactic regulations of codeswitching have been frequently studied and enriched through the study of morphosyntactic combinations of various languages with different typologies. A fruitful language pairing for studying codeswitching between is the suffix-dominated agglutinative Turkish, and the more simplistic/fusional lingua franca English, which this essay aims to investigate on morpho-syntactic grounds. The structural properties of codeswitching outlined by Poplack (1980) with the Free Morpheme and Equivalence Constraint, and later by Myers-Scotton's (2006) Matrix Language Frame (MFL) and 4M Model provide theoretical patterns for such bilingual combinations, with ample research testing their findings. Not enough attention, however, has been paid within the literature to agglutination and inflectional patterns of codeswitched words and phrases within this pairing. Therefore, this paper aims to address this gap in research by studying codeswitching patterns of bilingual speakers of Cypriot Turkish and English.

In her research on codeswitching habits of Turkish-English bilinguals in New York, Koban's (2013) findings avoid evaluation of the agglutination patterns of codeswitching by only studying the differentiation between intra- and inter-sentential switches, with the intra-sentential switches being

categorised into single-word and double-word categories depending on how many L2 words the speakers preferred to embed into their L1 in each utterance. Similarly, although Kemaloğlu-Er (2018) highlighted postpositional Turkish structures appearing within prepositional English forms, disproving Poplack's Free Morpheme Constraint that suggests bound morphemes like affixes cannot be embedded into another language, the researcher mentions very little about bilingual constructions that form complex multilingual morpheme/word clusters, going against the Free Morpheme Constraint. Turkish being a highly agglutinative language, 'nonce words' (Poplack, 1980) or embedded words are likely to receive multiple inflections even before being phonologically and morphologically integrated into the Matrix Language Frame (Myers-Scotton, 2006), which is the dominant model of language into which embeddings from a L2 are inserted. Therefore, while it is possible to come across English embedded words in their bare forms within Turkish phrases, they might also receive complex inflections from the Matrix language, under Myers-Scotton's 4M-model.

Moreover, in the case of embedded verbs from English into Turkish, the literature shows either a noun-to-verb agglutination process between the two languages, or a bilingual periphrasis function that combines the embedded verb/island with an auxiliary verb or catena from the Matrix structure, as further demonstrated later in Findings. Boumans (2007) points these two phenomena out, positing that the latter form can be regarded as a marker of intense language contact. Interestingly, Boumans' cross-linguistic analysis of these periphrastic formations provides a comparison of Cypriot Greek's high use of periphrasis in codeswitching and borrowings against Mainland Greek's use, stating that 'in 1960 Cyprus gained independence [as a British colony] and (Modern Standard) Greek became the language of administration. The English language remained influential through the tourism industry and the large international community on the island' (2007, p.300). While the researcher uses this in relation to intense language contact, Boumans fails to consider that the official languages of the Republic of Cyprus are Greek, along with Turkish and English. The intense contact formations, therefore, should be equally likely to appear in the island's second native language, which this essay's Turkish-speaking Cypriot sample, who speak Cypriot dialects of Turkish, and are all from backgrounds that value English's usefulness in inter- and intra-communal communications, will provide further evidence on. With these considered, the main questions of research emerge as:

1. Do English embedded lexemes receive multiple or complex Turkish inflections?
2. Do codeswitched Turkish VPs with English verbs occur more with periphrasis than inflections (perhaps due to intense language contact between Cypriot Turkish and English)?

Using the findings of Boumans and Myers-Scotton's updated MFL, we can hypothesise that the answers to both questions will be more on the affirmative side, given the linguistic characteristics of Turkish as a matrix language. While testing this speculation, this essay also aims to consider these questions along with the data produced with consideration of participant language attitudes towards codeswitching, for more wholistic speaker profiles.

2 Methods

To study these within limitations of time and brevity, the study was conducted with four participants chosen using a purposive sampling method. The specific speaker characteristics sought for are further discussed in the following section. The design of the study consisted of two one-hour semi-structured conversation sessions moderated by the researcher, with two separate, randomly-assigned speaker pairs. Before the sessions, all participants signed consent forms, and were given an overview of their right to

withdraw, with any arising questions answered by the researcher without any leading answers. The pairs in the sessions were allocated randomly, but since all speakers were picked from the same broad social group in Cyprus, they all had previous connections with each other. This assured that the participants were quicker to start a relaxed conversation, thus providing natural instances of codeswitching. As emphasised by Koban's (2013) design, the research tried to capture codeswitching between Cypriot Turkish-English bilinguals from the same social background as an unmarked speech variety that occurs organically within this wider community of Turkish-speaking Cypriot young-adults. Hence, although the researcher was an active listener within the conversations, only a few open-ended questions were provided to generate some discussions about language attitudes. This study design was also observed in Eversteijn's (2011) study on Turkish-Dutch bilingual teenagers, who highlighted the advantages of open-ended questions in groups of 2-4 participants, suggesting that 'informants started to compare and complement each other's answers.' Therefore, while the natural conversation of friends and organic occurrences of codeswitching were prioritised, a finer social profile of the speakers were created considering their discussions to better understand the function of codeswitching within their speech community.

The two sessions took place over video chat to stimulate a friendly online conversation space, generating about two hours of data. The first 45 minutes of each conversation were studied, and all occurrences of codeswitching, along with instances of syntactic mishaps, incoherence, or ungrammaticality caused by a mismatch between the two competing language structures, were transcribed and tabulated by the researcher. During transcriptions, the codeswitching tokens were separated into groups of intra-sentential, inter-sentential, and an additional category of 'extra-sentential' for tag interjections after Poplack (1980). While the intra-sentential switches were counted as L2 embeddings on single word or NP levels, the inter-sentential group consisted of CP-level switches. The third category was made to differentiate the more solidified forms of phrasal codeswitches that were referential to Anglophone internet culture such as well-known quotes from Vine, Tiktok, or collocations. Further distinctions were made between 'intra-clausal' (Myers-Scotton, 2006) morphological embeddings to look at the number of suffixes they received; and agglutinated/non-agglutinated VP structures to look at the suffixes accompanying the embedded verbs or catenae used for periphrasis, to help with coding. The participants were informed that this research was focused on codeswitching habits within their community, which was discussed in conversation amongst them, but were not given any definite cues as to what aspects of codeswitching were being studied and codified. The participants were allowed to start their sentences in any language they preferred throughout the session, and the moderator did not prompt the use of a specific language during the conversations. Therefore, while most recorded instances of codeswitching were English embedded word and islands within Cypriot Turkish matrix clauses, some instances such as reported speech codeswitches were with sentences that were formed in the English matrix frame.

3 Sample

As mentioned, the sample of this study were all chosen because they were from the same social circle of Turkish-speaking Cypriot teenagers in their early twenties. The intensity of language contact that can explain their codeswitching patterns is observed through their reports of multilingual living. All speakers had Turkish-speaking Cypriot parents with varying degrees of bilingualism within the immediate family. Three speakers, coded *I*, *B*, and *C*, received their secondary education in a private English school for both Greek-speaking and Turkish-speaking Cypriots, where the education was predominantly in English. One speaker, coded *A*, went to a Turkish-English grammar school, where Turkish was relatively used more frequently in administration and classrooms. While the former group report higher usages of

codeswitching within the Turkish-speaking Cypriot (TsC) youth at school, *A* states only a few people frequently codeswitched during conversations outside classrooms. More importantly, all participants are currently pursuing their higher education within the UK and have an advanced English proficiency from utilising it in their social, educational, and professional lives. Naturally, codeswitching is more common in speech amongst people from the same community who also study in the UK.

Moreover, the language attitudes measured by the participants' conversations springing from open-ended questions suggest that codeswitching within this community is a matter of comfort and 'unmarkedness' (Myers-Scotton, 1983). All speakers recognise that codeswitching outside of their community's discourses serves to create emphasis and acknowledge the difference of other bilingual varieties such as those of the Cypriot diasporas from their form of codeswitching arising from years of English education and mass media consumption, which suggests the forms studied could be more individualistic. The speakers emphasised that they explicitly codeswitch with a specific friend group, and the importance of (specifically Cypriot) Turkish for their social identities, highlighting their feeling of responsibility to keep their L1 'alive'. While the following discussions consider language contact and proficiency in both languages as an explanation for the phenomena studied, future researchers might want to conduct a more impressionistic study by focusing on other sociolinguistic factors such as age, class, socio-political identity development, etc. in relation to codeswitching in Cyprus.

4 Findings

Within 90 minutes of speech recorded between *B&I* and *A&C*, 143 tokens of codeswitching were collected. Five of these instances were ruled out from the following analysis for being syntactic errors in monolingual sentences (direct translations of phrases, wrong word order, etc.), but recorded for signalling non-selective activation. Of the total amount, 78% of codeswitches were single or phrase-level nouns, adjectives, tag insertions, and TP/CP level inter-sententials. The breakdown of the classifications of these codeswitches are represented in Figure 1.

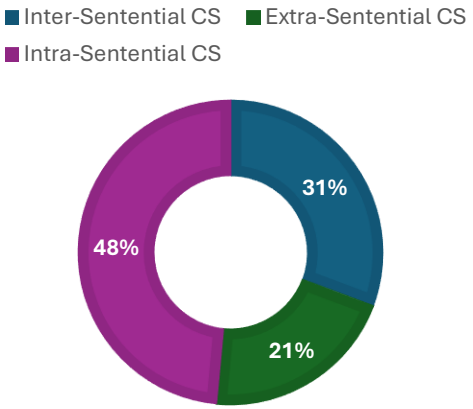


Figure 1: *Pie chart depicting the percentages of the classifications of recorded codeswitching tokens (excluding VP- level embedding), N=111.*

Most of the non-VP-level codeswitches were "intra-clausal" (Myers-Scotton, 2006), with extra- and intra-sententials occurring on word or phrase-levels. The majority of the extra-sentential codeswitches recorded were phrase insertions such as '*oh my God*, o zaman iyi para alın' (meaning *'oh my God, then you are getting good money'*), and word-level fillers like 'literally', 'obviously'. The majority of the recorded inter-sententials were English embedded phrases within a Turkish Matrix Frame and were marked by a

complementiser discourse marker (functioning as a content morpheme (Myers-Scotton, 2006)) such as 'yani' (*'so'*) or 'like', (see Example (1), spoken by *C*), or a slight pause.

(1) Turkish-English
*yani **it's seen as like demeaning**; gendi kültür -ün -ü bil -me -n*
so <it's seen as like demeaning>, own culture -2SG.POSS -ACC know -NEG -2SG
'I mean **it's seen as like demeaning**; you do not know your own culture.'

Importantly, of the recorded intra-sentential codeswitches, 38 tokens were lexical borrowings, and 16 were NP structures like "government building", "advanced research centre", etc. These were the tokens observed to study the complexity of agglutination and number of affixes, producing Figure 2.

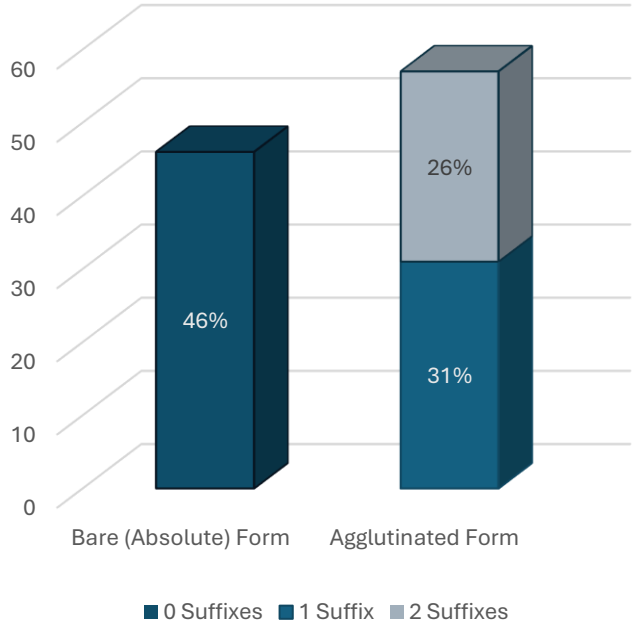


Figure 2: *Percentages of bare and agglutinated embedded words and Noun Phrases, differentiating the number of added suffixes, N=54.*

It was found that suffixation of embedded words happens more frequently, however a standard pattern for single, double, or more suffixation was not observed. In embedded lexemes and NPs, the most suffixes received were two. The most used suffixes were the dative (-e/-a such as in 'grammar-a', '*to the grammar'*) and accusative cases (-ı/-i/-u/-ü such as in 'o line-ı okuduğumda', '*when I read that line'*), along with pluralisation (-ler/-lar in 'sexuality-ler', '*sexualities*'). On double suffixed words, these were usually followed by a possessive (see Ex. 2, spoken by *A*).

(2) Turkish
*el gol hareket -ler -im ə:: **expression-lar-un** da*
hand arm movement -PL -1SG.POSS ə:: <expression>-PL-1SG.POSS too
değiş -me –(y)e başla-r.
change -INF -DAT start-PRS
'my hand-arm movements, uhh, my **expression**s start to change too.'

In the case of verbs and VP-level codeswitches that formed 18% of the total score of tokens, agglutination happened with up to 4 suffixes. These tokens that are not represented in Figure 1 show that the bilingual VPs formed with agglutination used either the N-to-V suffix '-le(mek)', *'(to) do*', or '-dı', '*to be.(PST)'*, along with regular inflectional morphemes showing person, tense, and negation, such as in Example 3 uttered by *I*:

(3) Turkish
submit *-le -di –(y)di -m*
<submit> -do -PST -IPVF -1SG

These forms of VPs only formed 31% of the bilingual verb and VP-level formations, with the majority being formed with periphrasis using Turkish auxiliary catenae, or 'light verbs' (Alexiadou, 2011) such as 'yap(mak)', *'(to) make*', 'et(mek)', *'(to) do'*, or 'ol(mak)', '*(to) be*' (see Fig. 3). In periphrastic forms, the embedded verbs retained their bare form, while the catenae were agglutinated, and as depicted, periphrasis was preferred more when embedding verbs into speech. Most bilingual VP creations preferred periphrasis with the light verb meaning *to make*, as exemplified by the gloss in Example 4 uttered by *B.*

(4) Turkish
Hayir, düşün-dü -m ve ***relate yap -dı -m****!*
No think -PST -1SG and <**relate**> make -PST -1SG
'No, I thought about it, and **I relate to it**!'

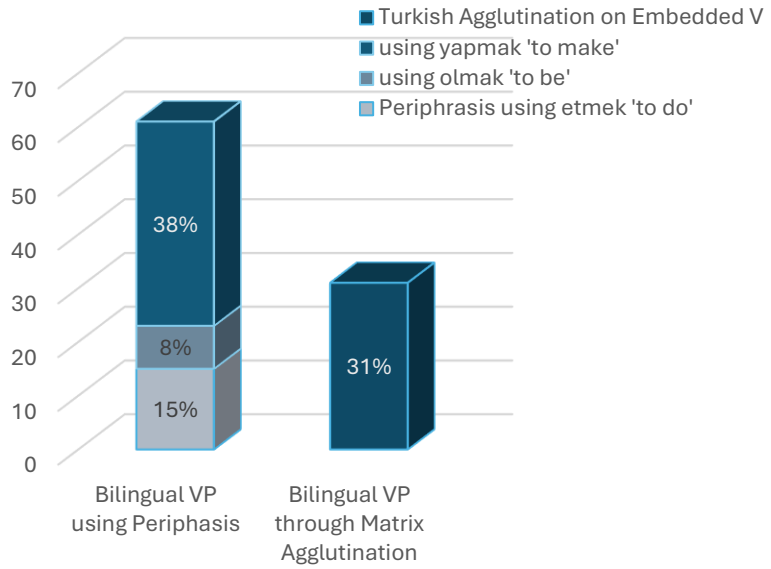


Figure 3: *Percentages of Verb Phrases formulated with embedded verbs and agglutination or periphrasis, with the additional distinction of auxiliary verbs used in the latter, N=26.*

5 Discussion

Recalling the two research questions as:

1. Do English embedded lexemes receive multiple or complex Turkish inflections? and
2. Do codeswitched Turkish VPs with English verbs occur more with periphrasis than inflections?,

the results seem to provide mostly affirmative answers. Although not enough data is gathered to pass a rigid judgement about the 'complexity', it is seen that embedded words do get agglutinated under the Matrix Language's regulations instead of remaining in simple (bare) forms, proving that the Matrix Language descends even into the morphological units of the Embedded Language (Myers-Scotton, 2006). This observed infiltration reflects the much-discussed dilemma between embeddings and borrowings, which tread the line of inseparability due to their high-level integration (Poplack & Dion, 2012; Myers-Scotton, 2006). The high number of extrasentential tags and multiple-worded collocational intra-/inter-sentential switches (see Fig. 1) also affirm Myers-Scotton's theory about the larger fragments or islands of Embedded Language remaining intact to avoid the constraints of the Matrix Language (2006), as their structures were mostly retained due to speakers syntactically manoeuvring around them, as evident in Example 1. Moreover, the distinction of extra-sententials during coding highlighted the effect of social media, and networks utilising an English sociolect which emphasises the importance of referential internet culture, on the codeswitching habits of the speakers. Although Melado & Lignos (2022) state that abbreviations like "OMG" were particles from a not-necessarily-bilingual 'internet dialect', the sample's codeswitches definitely accentuated these utterances and hinted at a sociolinguistic purpose for this form of codeswitching within this community of speakers. Codeswitching in general, whether Turkish-to-English or vice versa, seems to have an additional sociolinguistic role within the bilingual Cypriot community, as speakers reported that they also inserted Cypriot Turkish into their English as an act of identity building and preservation.

In terms of intra-sentential codeswitches being preferred more, the findings coincide with Koban's (2013) hypothesis that advanced bilinguals are more likely to codeswitch intrasententially, since the sample were proficient in both languages from an early age (see Sample). However, the groups' familiarity and recognition of each other's proficiencies did not influence the complexity of suffixation of embedded nouns, adjectives, and NP's, since embeddings did not receive that many early or late system morphemes, as predicted by Myers-Scotton's 4M-Model (2006). Excluding verbs and VPs, embedded words received a maximum of 2 suffixes, which, when compared to matrix noun suffixations (see 'hareket-lerim' and 'expression-lar-ım' in Ex. 2), does not appear particularly dense. Here, we also see Kemaloğlu-Er's (2018) results affirmed, and Poplack's Free Morpheme Constraint (1980) debunked once again, as speakers used English free morphemes with Turkish bound morphemes regularly.

In relation to periphrasis and intense language contact, the results were mostly supportive of Boumans' (2007) statements. Although Boumans suggested that Turkic languages use periphrasis in 'virtually all language contact situations', some noteworthy forms of agglutinative bilingual verb-formation were also observed. His assertion concerning Cypriot Greek either applying complete morphological integration to embedded verbs with 'ML axes', or using periphrasis with auxiliary 'light verbs' (Alexiadou, 2011; see Ex. 4) to be inflected instead of the embedded verb, seems to apply more correctly to Cypriot Turkish as well. Considering the sample as representative for the community of young Turkish-speaking Cypriots who use English inter- and intra-communally, the hypothesis that intense language contact results in more observations of bilingual periphrasis is supported to an extent (see Fig. 3). Similarly, the structures of verb-creating agglutination were preceded by Alexiadou (2011), who stated that embedded verbs were only through 'indirect insertion' after a N-to-V class-changing suffix was applied first. This is exemplified by the use of the highly productive verb-forming suffix '-le(mek)' as a

bordering affix, allowing further inflection only after 'changing the category' of the embedded verb (see Ex. 3). These forms of bilingual VP-creation signify both the advanced proficiencies of the speakers in both languages, along with the intensity of language contact and codeswitching within their daily communications. Furthermore, the choices of embedded lexemes suggest that codeswitching resulted from certain semantic fields being acquired only in one language. All participants noted the absence of some subject-specific terms in Turkish, and English academic jargon's irreplaceable 'naturalness'.

Overall, through studying the morphosyntax of codeswitching of Cypriot Turkish-English bilingual young-adults, the study yields intriguing data in support of a positive correlation between bilingual periphrasis and intense language contact as posited by Boumans (2007), and interesting agglutination patterns observed on embedded words, in line with Myers-Scotton's 4M-Model (2006). Within this social group, positive language attitudes towards the phenomenon are observed within the sample, reflecting that English embeddings are not seen as deformative, but a sociolinguistic identity-building tool that should be researched further, with the speakers using the two languages to fill in the semantic gaps.

Finally, while the design of the study might have caused biases such as an observer's paradox, organic instances of codeswitching were captured by the researcher, that represent the natural intra-communal utilisation of codeswitching within the young community of Cypriot Turkish-English bilinguals.

6 References

Alexiadou, A. (2011). Remarks on the morpho-syntax of code-switching. *Proceedings of the 9th international Conference on Greek Linguistics*, 44–55.

Boumans, L. (2007). The periphrastic bilingual verb construction as a marker of intense language contact. Evidence from Greek, Portuguese and Maghribian Arabic. In E. Ditters & H. Motzki (Eds.), *Approaches to Arabic linguistics* (pp. 291–311). Leiden: Brill.

Eversteijn, N. (2011). *"All at once": Language choice and codeswitching by Turkish-Dutch teenagers.*

Kemaloğlu-Er, E. (2018). Patterns of Intrasentential Code-Switching in Turkish-English Bilingual Discourse: Testing the Free Morpheme and the Equivalence Constraint. *Artıbilim: Adana Science and Technology University Journal of Social Science 2018*, 1(2), 35–45. https://dergipark.org.tr/tr/download/article-file/612458

Koban, D. (2013). Intra-sentential and Inter-sentential Code-switching in Turkish-English Bilinguals in New York City, U.S. *Procedia - Social and Behavioral Sciences*, 70, 1174–1179. Science Direct. https://doi.org/10.1016/j.sbspro.2013.01.173

Mellado, E. A., & Lignos, C. (2022). *Borrowing or Codeswitching? Annotating for Finer-Grained Distinctions in Language Mixing.* arXiv preprint arXiv:2206.04973.

Myers-Scotton, C. (1983). The negotiation of identities in conversation: a theory of markedness and code choice. *International Journal of the sociology of Language* 1983(44), 115-136. https://doi.org/10.1515/ijsl.1983.44.115

Myers-Scotton, C. (2006). *Multiple Voices: An Introduction to Bilingualism*. Chichester: Wiley-Blackwell.

Poplack, S. & Dion, N. (2012). Myths and facts about loanword development. *Language variation and change*, 24(3): 279-315.

Poplack, S. (1980). Sometimes I'll start a sentence in Spanish Y TERMINO EN ESPAÑOL: toward a typology of code-switching. *Linguistics*, 18(7-8). https://doi.org/10.1515/ling.1980.18.7-8.581

Sound Symbolism and Dimensionality

Lydia Wiernik
University of Edinburgh

Abstract: The bouba/kiki paradigm is a product of sound symbolism, the linguistic phenomenon of meaning/sound associations. First documented in 1929, the paradigm has become the focus of numerous studies. However, despite being so well-attested, there is no single, accepted explanation for its existence. In the present study, I take D'Onofrio's (2014) position, which suggests the influence of vowel frontness and backness on the paradigm. D'Onofrio (2014) is novel in that it departs from the abstract, two-dimensional stimuli characteristic of previous studies instead uses familiar, three-dimensional objects as stimuli. After providing a brief overview of D'Onofrio (2014) and other literature, I introduce an online experiment designed to take our knowledge of this paradigm further. As in D'Onofrio (2014), I use three-dimensional stimuli, but introduce another element: physical movement via GIFs. This places the paradigm further into the real world by simulating physical interaction. My hypotheses are twofold: first, that form-meaning association is significantly influenced by three-dimensionality; and second, that the perceived 'sharpness' or 'roundness' of the stimuli is boosted by the three-dimensionality of these properties. It is important to note, however, that no data has been collected regarding this experiment. This writeup only summarises the aims of the experiment created and introduces a proposal for its use, and does not make any data-driven claims about these hypotheses. This proposal and the experiment serve only as a suggestion for the direction of future work.

Keywords: sound symbolism; dimensionality; vowel quality; phonetics; phonology; cognition

---

1 Background

Sound symbolism is the linguistic phenomenon of meaning/sound association. The bouba/kiki effect is one of its most well-known paradigms, first documented by Wolfgang Köhler (1929) and replicated ad infinitum, notably by Ramachandran and Hubbard (2001) who coined the labels 'bouba' and 'kiki'. In this paradigm, rounded objects are consistently paired with words also perceived to be round (like 'bouba') and spiked objects are paired with those perceived to be sharper (like 'kiki'). There have been various explanations for this; Ramachandran and Hubbard (2001) suggested influence of the mouth's roundness when producing the word 'bouba' and tautness when producing 'kiki.' Other studies have considered the role of different consonants, vowel quality, and even vowel quantity (Bross 2018).

In a novel approach, D'Onofrio (2014) considers vowel frontness and backness. Additionally, she departs from previous studies (Ramachandran and Hubbard (2001); Maurer (2006)) which have only used abstract, two-dimensional shapes as stimuli, and instead uses familiar, three-dimensional objects. D'Onofrio (2014) finds that vowel frontness and backness are significantly influenced by the dimensionality of an object; perceived roundness was associated with back vowels and perceived spikiness with front vowels. This expands the paradigm's reach, situating it in the real world.

In this writeup, I continue D'Onofrio's exploration and investigate the extent of the paradigm's real-world applications by proposing an experiment which introduces three-dimensional GIFs as stimuli, simulating tactile interaction. I maintain that investigating language users' conceptions of dimensionality, especially three dimensionality, is the next step in understanding the cognitive processes behind sound symbolic associations.

2 Motivation

Like D'Onofrio (2014), gaps in the literature motivated my decision to use three dimensional objects. Many studies have only tested dimensionality relative to two dimensions. For example, Bankieris et al. (2015) and Cuskley et al. (2019) found sound-symbolic associations with vowel category were weakest when the domain of meaning was directional (up/down) or spatial. However, both these domains are situated on the X/Y axis; exploring the introduction of a Z-axis – three dimensionality – may elicit different results.

Indeed, D'Onofrio's (2014) findings suggest that the Z-axis plays an untapped role in sound-symbolic associations. Notably, D'Onofrio (2014) found that stimuli with three-dimensional conceptions of 'roundness' were even *more* significant predictors of vowel backness than stimuli with two-dimensional 'roundness.' The newfound significance of three-dimensionality has implications for the presence of the bouba/kiki paradigm and sound symbolism in the real world. It suggests that these phenomena may be more common cognitive processes than previously thought, perhaps often activated in our daily lives where we constantly interact with three-dimensional objects.

This finding abstractly contradicts Bankieris et al. (2015) and Cuskley et al.'s (2019) findings that X/Y axis dimensionality did not correspond positively to sound symbolic associations. However, it may be that dimensionality must be form-related, not meaning-related, to be significant, or that X/Y axis dimensionality is not as salient a predictor as Z-axis.

2.1 Departing from D'Onofrio (2014)

The present experiment puts D'Onofrio's (2014) preliminary findings on three dimensionality at the forefront, expanding the number of stimuli used and focusing directly on dimensionality's role, rather than observing it as a secondary finding. Further, though D'Onofrio (2014) found three-dimensionality a significant predictor, this result only applied to three-dimensional *roundness*. Therefore, in this experiment, I have used the same procedure for spikiness to determine if either more data or learning bring out the same effect.

3 Research questions

Primarily, my experiment investigates whether using interactive 3D rendered stimuli influences the vowel quality of participants' object descriptions.

The combination of sound symbolism and iterated learning also allows for an examination of their interaction, that is, whether one takes precedence over another; will the bouba/kiki paradigm (or broader facets of sound symbolism) override learning?

My predictions are as follows:

1. form-meaning association is significantly influenced by three-dimensionality
    a. and if it is, this will be shown by convergence to the predicted vowel quality (either backness or frontness) when paired with the opposite one
2. the perceived 'sharpness' or 'roundness' of the stimuli is boosted by the three-dimensionality of these properties
    a. and if it is, the vowel association effects will be boosted

4 Experimental design and set-up

4.1 Experiment technicality and functionality

The present experiment was coded using JavaScript and the JsPsych behavioural experiment framework and runs on any desktop browser. The code is not yet open source accessible as this work is ongoing.

The experiment tests a corpus of 18 form-meaning pairs of real-world objects and 'alien' labels. The experiment begins with several instruction and consent screens, telling participants they will be learning a created 'alien' language. The instructions ensure the participant has configured their audio correctly, so as not to lose any data unnecessarily.

Participants then enter the learning phase and are exposed to a randomised subset of ten object-label pairs. They then move on to the production phase, where they must record their own audio labels for the entire corpus of 18 pairs. The subset in the learning phase simulates a bottleneck on transmission in order to observe convergence to a certain form.

4.2 Visual and audio stimuli

The visual stimuli are divided evenly between three spiky and three round forms with three variations on dimensionality each: a 2D image, a 3D static image, and a 3D GIF. Within the round or spiky groupings, two are classically round or spiky, and one is in the middle, termed the 'blend.' The blend has been added to investigate which of the two form qualities, round or spiky, is more salient by observing which prevails during the labelling process. Since D'Onofrio did not find three-dimensionality significant for spikiness, my introduction of the blend secondarily investigates the strength of (or discrepancy between) the form qualities themselves.

All visual stimuli are familiar objects in order to simulate as close as possible a real-world experience. See Appendix A for all visual stimuli. The auditory stimuli are the same as in D'Onofrio (2014), given below in Figure 1. The original audio files were not made available for reuse, and were re-recorded for the purposes of this experiment.

Table 3. Pool of martian words for the real-world object task.

Condition	Front vowels		Back vowels	
	/i/	/e/	/u/	/ɑ/
1	/pimə/	/pemə/	/pumə/	/pɑmə/
2	/bimə/	/bemə/	/bumə/	/bɑmə/
3	/timə/	/temə/	/tumə/	/tɑmə/
4	/dimə/	/demə/	/dumə/	/dɑmə/
5	/kimə/	/kemə/	/kumə/	/kɑmə/
6	/gimə/	/gemə/	/gumə/	/gɑmə/

Figure 1: *Original audio stimuli for generation 0.*

4.3 Reasoning for dimensionality variations

Splitting dimensionality into three variations, each more physically dimensional than the previous, will allow me to total the proportion of vowel backness for each and thus more accurately measure how dimensionality influences results.

This approach takes from D'Onofrio (2014), where she refines the original bouba/kiki experiment by individually testing possible influencing features (such as vowel height or mouth shape) to see what each brings to sound symbolic associations. D'Onofrio essentially reverse engineers the paradigm, working backwards to determine which features are responsible for which associations. This helps to more fully understand the cognitive process; I believe this is an equally relevant tactic for the present experiment.

For each object, the 2D and 3D static dimensionality types are matched with the 'correct' vowel for their shape. That is, round objects are matched with back vowels; spiky objects are matched with front vowels. The 3D GIF, however, is matched with the 'incorrect' vowel, serving as the critical trial. This is visualised in Figure 2.

This mismatch tests my primary prediction outlined in section 2. The critical trial tests whether participants map the 'correct' vowel quality to a 3D stimulus, even when it is labelled with the 'incorrect' vowel quality. This measures whether participants are significantly influenced by dimensionality – labelling the 3D GIF 'correctly' regardless of training – or if through learning, they converge toward a different vowel.

Shape	Object	Dimensionality type	Vowel type
Round	Wheel	2D 3D static	Back
		3D gif	Front
	Glass	2D 3D static	Back
		3D gif	Front
	Succulent	2D 3D static	Back
		3D gif	Front
Spiky	Knife	2D 3D static	Front
		3D gif	Back
	Ladder	2D 3D static	Front
		3D gif	Back
	Christmas tree	2D 3D static	Front
		3D gif	Back

Figure 2: *The initial language in the experiment. The yellow shaded conditions represent critical trials.*

A result supporting my primary prediction would find, for example, the proportion of back vowel productions for a round 3D GIF higher than in any other condition, especially *over* the 3D static image. Since movement (simulated interaction) would be the only difference between 3D static and 3D GIF stimuli, this result would situate this facet of dimensionality – which is most connected to our real-world experiences – as a significant domain influencing cognition.

An especially significant result would find participants converging towards a certain vowel, even if it wasn't present in the data (e.g., it had disappeared over generations of learning). This would situate the power of the association above the effects of learning.

5 Scope for further work

There is wide scope for studies building on this experiment, particularly those incorporating the cognitive phenomenon of synaesthesia. The literature is rife with discussion on the connection between sound-symbolism and synaesthesia: Bankieris & Simner (2015) note they both have been linked to the left superior parietal cortex; Bankieris & Simner (2015) and Cuskley et al. (2019) suggest that synaesthesia is a latent cognitive function, activated in some but possible in all. Given this neurological linkage, it would be apt to test whether dimensionality would bring to the fore these mechanisms, and if invoking one of sound symbolism and synaesthesia in turn invokes the other. If these phenomena share an apparatus, further study could provide insight into the cognition behind both, and perhaps their interaction with each other.

6 Conclusion

The present proposed experiment fills a gap in the current literature, bringing the bouba/kiki paradigm further into the real world and allowing for a more adequate understanding of the cognitive processes behind sound symbolism. A study such as this may help unravel our cognitive perceptions and conceptualizations of dimensionality, as well as allowing a deeper understanding of how sound-symbolism and dimensionality function – individually, together, and under the influence of iterated learning.

This study is ongoing; future work will focus on data collection and analysis, accompanied by a second write-up detailing results.

7 References

Bankieris, K., & Simner, J. (2015). What is the link between synaesthesia and sound symbolism? *Cognition*, 136, 186–195. https://doi.org/10.1016/j.cognition.2014.11.013

Bike Wheel. (n.d.). Envato Elements. Retrieved December 3, 2022, from https://elements.envato.com/bike-wheel-TW7CY3

Bike Wheel Stock Illustrations – 76,547 Bike Wheel Stock Illustrations, Vectors & Clipart - Dreamstime. (n.d.). Bike Wheel Stock Illustrations – 76,547 Bike Wheel Stock Illustrations, Vectors & Clipart - Dreamstime. Retrieved December 3, 2022, from https://www.dreamstime.com/illustration/bike-wheel.html

Bross, F. (2018, January 1). *Cognitive associations between vowel length and object size: A new feature contributing to a bouba/kiki effect*. Conference on Phonetics & Phonology in German-Speaking Countries. https://www.researchgate.net/publication/323749384_Cognitive_associations_between_vowel_length_and_object_size_A_new_feature_contributing_to_a_boubakiki_effect

Cuskley, C., Dingemanse, M., Kirby, S., & van Leeuwen, T. M. (2019). Cross-modal associations and synesthesia: Categorical perception and structure in vowel–color mappings in a large online sample. *Behavior Research Methods*, 51(4), 1651–1675. https://doi.org/10.3758/s13428-019-01203-7

D'Onofrio, A. (2014). Phonetic Detail and Dimensionality in Sound-shape Correspondences: Refining the Bouba-Kiki Paradigm. Language and Speech, 57(3), 367–393. https://doi.org/10.1177/0023830913507694

Douglas Fir Tree. (n.d.). Envato Elements. Retrieved December 3, 2022, from https://elements.envato.com/douglas-fir-tree-PL4LR3

Empty glass (line drawing) - Stock Illustration [81106288] - PIXTA. (n.d.). @PixtaStock. https://www.pixtastock.com/illustration/81106288

FAVPNG.com. (n.d.-a). *Drawing Christmas Tree Line Art - PNG - Download Free*. FAVPNG.com. Retrieved December 3, 2022, from https://favpng.com/png_view/christmas-drawing-christmas-tree-line-art-png/qMCYpm3u

FAVPNG.com. (n.d.-b). *Drawing Clip Art - PNG - Download Free*. FAVPNG.com. Retrieved December 3, 2022, from https://favpng.com/png_view/ladders-drawing-clip-art-png/s7WW9jn5

Hand-drawn contour of succulents and other houseplants. Vector... (n.d.). IStock. Retrieved December 3, 2022, from https://www.istockphoto.com/vector/succulents-draft-set-gm1139874500-304820764

Kitchen Knife. (n.d.). Envato Elements. Retrieved December 3, 2022, from https://elements.envato.com/kitchen-knife-U7AAYTF

Kohler, W. (1929). *Gestalt psychology, an introduction to new concepts in modern psychology*. New York: Liveright.

Ladder. (n.d.). Envato Elements. Retrieved December 3, 2022, from https://elements.envato.com/ladder-K3RPZDP

Maurer, Daphne; Pathman, Thanujeni & Mondloch, Catherine J. (2006). "The shape of boubas: Sound-shape correspondences in toddlers and adults" (PDF).

Pint Glass. (n.d.). Envato Elements. Retrieved December 3, 2022, from https://elements.envato.com/pint-glass-6K5JCR

Premium Vector | A large handdrawn knife doodle kitchen accessories vector. (n.d.). Freepik. Retrieved December 3, 2022, from https://www.freepik.com/premium-vector/large-handdrawn-knife-doodle-kitchen-accessories-vector_28912492.htm

Ramachandran, V. S., & Hubbard, E. M. (2001). Synaesthesia —A Window Into Perception, Thought and Language. *Journal of Consciousness Studies*, 8(12), 3–34. http://cbc.ucsd.edu/pdf/Synaesthesia%20-%20JCS.pdf

Succulent. (n.d.). Envato Elements. Retrieved December 3, 2022, from https://elements.envato.com/succulent-EMHBT7

8 Appendix One: Visual stimuli

Object	*2D*	*3D static*	*3D GIF*
Kitchen knife (classic)			
Ladder (classic)			
Christmas tree (blend)			

Figure A1: *Visual stimuli for "spikiness." Sources cited in references. GIFs I have made myself. Note that GIF format is unsupported by Word; in motion, the GIFs rotate 360° horizontally and vertically.*

Object	*2D*	*3D static*	*3D GIF*
Glass (classic)			
Wheel (classic)			
Succulent (blend)			

Figure A2: *Visual stimuli for "roundness." Sources identical to those of Figure A1.*

9 Appendix Two: Previous experiment runs

object	label
images/2d_wheel_1.png	sounds/puma.wav
images/3d_wheel.png	sounds/pama.wav
images/wheel.gif	sounds/pima.wav
images/2d_glass_1.png	sounds/buma.wav
images/3d_glass.png	sounds/bama.wav
images/glass.gif	sounds/bima.wav
images/2d_succulent_1.png	sounds/tuma.wav
images/3d_succulent.png	sounds/tama.wav
images/succulent.gif	sounds/tima.wav
images/2d_knife_1.png	sounds/dima.wav
images/3d_knife.png	sounds/dema.wav
images/knife.gif	sounds/duma.wav
images/2d_ladder_1.png	sounds/kima.wav
images/3d_ladder.png	sounds/kema.wav
images/ladder.gif	sounds/kuma.wav
images/2d_christmas_tree_1.png	sounds/gima.wav
images/3d_christmas_tree.png	sounds/gema.wav
images/christmas_tree.gif	sounds/guma.wav

Figure A3: *Original input language, generation 0.*

participant_id	chain	generation	trial_index	block	time_elapsed	stimulus	sound
ozslte8u7k	1	1	7	observation	25583	images/succulent.gif	sounds/tima.wav
ozslte8u7k	1	1	8	observation	27354	images/2d_christmas_tree_1.png	sounds/gima.wav
ozslte8u7k	1	1	9	observation	28616	images/3d_christmas_tree.png	sounds/gema.wav
ozslte8u7k	1	1	10	observation	29820	images/glass.gif	sounds/bima.wav
ozslte8u7k	1	1	11	observation	30708	images/christmas_tree.gif	sounds/guma.wav
ozslte8u7k	1	1	12	observation	31263	images/3d_knife.png	sounds/dema.wav
ozslte8u7k	1	1	13	observation	32836	images/knife.gif	sounds/duma.wav
ozslte8u7k	1	1	14	observation	33254	images/3d_ladder.png	sounds/kema.wav
ozslte8u7k	1	1	15	observation	33697	images/2d_ladder_1.png	sounds/kima.wav
ozslte8u7k	1	1	16	observation	34147	images/ladder.gif	sounds/kuma.wav
ozslte8u7k	1	1	19	production	45903	images/glass.gif	sounds/ozslte8u7k_0.webm
ozslte8u7k	1	1	21	production	49403	images/2d_wheel_1.png	sounds/ozslte8u7k_1.webm
ozslte8u7k	1	1	23	production	52892	images/2d_glass_1.png	sounds/ozslte8u7k_2.webm
ozslte8u7k	1	1	25	production	58070	images/knife.gif	sounds/ozslte8u7k_3.webm
ozslte8u7k	1	1	27	production	63452	images/2d_succulent_1.png	sounds/ozslte8u7k_4.webm
ozslte8u7k	1	1	29	production	67460	images/3d_wheel.png	sounds/ozslte8u7k_5.webm
ozslte8u7k	1	1	31	production	74556	images/succulent.gif	sounds/ozslte8u7k_6.webm
ozslte8u7k	1	1	33	production	79447	images/3d_knife.png	sounds/ozslte8u7k_7.webm
ozslte8u7k	1	1	35	production	82984	images/christmas_tree.gif	sounds/ozslte8u7k_8.webm
ozslte8u7k	1	1	37	production	88050	images/3d_glass.png	sounds/ozslte8u7k_9.webm
ozslte8u7k	1	1	39	production	92817	images/ladder.gif	sounds/ozslte8u7k_10.webm
ozslte8u7k	1	1	41	production	95676	images/2d_knife_1.png	sounds/ozslte8u7k_11.webm
ozslte8u7k	1	1	43	production	100887	images/wheel.gif	sounds/ozslte8u7k_12.webm
ozslte8u7k	1	1	45	production	104698	images/3d_ladder.png	sounds/ozslte8u7k_13.webm
ozslte8u7k	1	1	47	production	108114	images/3d_christmas_tree.png	sounds/ozslte8u7k_14.webm
ozslte8u7k	1	1	49	production	111521	images/2d_ladder_1.png	sounds/ozslte8u7k_15.webm
ozslte8u7k	1	1	51	production	114235	images/2d_christmas_tree_1.png	sounds/ozslte8u7k_16.webm
ozslte8u7k	1	1	53	production	117888	images/3d_succulent.png	sounds/ozslte8u7k_17.webm

Figure A4: *Generation 1; the first participants' responses.*

participant_id	chain	generation	trial_index	block	time_elapsed	stimulus	sound
bamm7bvo5j	1	2	7	observation	106551	images/3d_glass.png	sounds/ozslte8u7k_9.webm
bamm7bvo5j	1	2	8	observation	108355	images/2d_succulent_1.png	sounds/ozslte8u7k_4.webm
bamm7bvo5j	1	2	9	observation	110140	images/wheel.gif	sounds/ozslte8u7k_12.webm
bamm7bvo5j	1	2	10	observation	111772	images/3d_christmas_tree.png	sounds/ozslte8u7k_14.webm
bamm7bvo5j	1	2	11	observation	113370	images/christmas_tree.gif	sounds/ozslte8u7k_8.webm
bamm7bvo5j	1	2	12	observation	114721	images/2d_glass_1.png	sounds/ozslte8u7k_2.webm
bamm7bvo5j	1	2	13	observation	116301	images/3d_knife.png	sounds/ozslte8u7k_7.webm
bamm7bvo5j	1	2	14	observation	117941	images/ladder.gif	sounds/ozslte8u7k_10.webm
bamm7bvo5j	1	2	15	observation	119660	images/3d_wheel.png	sounds/ozslte8u7k_5.webm
bamm7bvo5j	1	2	16	observation	121128	images/2d_knife_1.png	sounds/ozslte8u7k_11.webm
bamm7bvo5j	1	2	19	production	129566	images/knife.gif	sounds/bamm7bvo5j_0.webm
bamm7bvo5j	1	2	21	production	133223	images/3d_glass.png	sounds/bamm7bvo5j_1.webm
bamm7bvo5j	1	2	23	production	138965	images/3d_wheel.png	sounds/bamm7bvo5j_2.webm
bamm7bvo5j	1	2	25	production	146078	images/2d_glass_1.png	sounds/bamm7bvo5j_3.webm
bamm7bvo5j	1	2	27	production	150608	images/2d_succulent_1.png	sounds/bamm7bvo5j_4.webm
bamm7bvo5j	1	2	29	production	157116	images/3d_ladder.png	sounds/bamm7bvo5j_5.webm
bamm7bvo5j	1	2	31	production	159992	images/2d_christmas_tree_1.png	sounds/bamm7bvo5j_6.webm
bamm7bvo5j	1	2	33	production	167022	images/ladder.gif	sounds/bamm7bvo5j_7.webm
bamm7bvo5j	1	2	35	production	169332	images/2d_knife_1.png	sounds/bamm7bvo5j_8.webm
bamm7bvo5j	1	2	37	production	173699	images/3d_christmas_tree.png	sounds/bamm7bvo5j_9.webm
bamm7bvo5j	1	2	39	production	178903	images/wheel.gif	sounds/bamm7bvo5j_10.webm
bamm7bvo5j	1	2	41	production	182386	images/christmas_tree.gif	sounds/bamm7bvo5j_11.webm
bamm7bvo5j	1	2	43	production	186524	images/succulent.gif	sounds/bamm7bvo5j_12.webm
bamm7bvo5j	1	2	45	production	189607	images/3d_succulent.png	sounds/bamm7bvo5j_13.webm
bamm7bvo5j	1	2	47	production	195078	images/glass.gif	sounds/bamm7bvo5j_14.webm
bamm7bvo5j	1	2	49	production	198044	images/2d_wheel_1.png	sounds/bamm7bvo5j_15.webm
bamm7bvo5j	1	2	51	production	200921	images/2d_ladder_1.png	sounds/bamm7bvo5j_16.webm
bamm7bvo5j	1	2	53	production	203306	images/3d_knife.png	sounds/bamm7bvo5j_17.webm

Figure A5: *Generation 2; the participant has learned from the previous participant in Generation 1.*

10 Acknowledgements

The present experiment combines aspects of Beckner et al.'s (2017) iterated learning and Loy & Smith's (2020) confederate priming code. I am grateful to them both for the accessibility of their code and encouragement to repurpose and reformulate it.

"That's not what they said": Verbal Misalignment in Subtitle Processing and its Effects on Cognitive Load

Aaron Merriel
University of Edinburgh

Abstract. Subtitles are a useful tool for translating media and aiding in learning and accessibility, but the use of auto-generation software means that often subtitles are not accurate to what was actually said in videos. This current work builds on existing research on how the brain processes subtitles by investigating the effects of lexical errors on cognitive load. Using data collected in an eye tracking experiment, it was investigated how viewers distribute their attention when there is misalignment between the different forms of sensory input. The results show that lexical errors caused a significant increase in cognitive load (measured by fixation duration), which suggests that errors impact the consolidation of subtitle information. Along with this, lexical errors caused viewers to change their viewing behaviour patterns such as revisiting subtitle areas or not reading certain subtitles. It was also found that there is an effect of time on fixation count, dwell time, and the viewing behaviours which requires further investigation.

Keywords: subtitle processing; cognition; cognitive load; eye tracking

---

1 Introduction

During the COVID-19 pandemic, people were online more than ever before. Whether it was watching television, working from home, or attending classes online, people were existing in an increased number of digital spaces. This change caused many people to become aware of how different consuming digital content is to in-person interactions. An interesting facet of consuming digital content is how it has multiple streams of input people receive simultaneously. This information can include spoken language, visual imagery, written language, and other auditory components such as music. While all of these can be experienced in in-person environments, digital content usually has all the streams of input coming from a single source.

Such differences raise questions about how the brain copes with these multiple streams of information simultaneously. One example of these simultaneous streams of input is the presence of audio and textual input in videos with subtitles. Subtitles, sometimes called captions, are any form of written text which appears overtop a primarily visual piece of media such as movies, YouTube videos, or online lectures. People consume subtitles in a variety of forms, whether subtitles are translating the dialogue into another language (inter-lingual), transcribing the dialogue spoken in the same language (intra-lingual), or supplying additional information to the audience about the situation in the film (auxiliary). In terms of multisensory information, intra-lingual subtitles are interesting because they are presenting the same linguistic information[2] in two different types of input: audio-verbal and audio-visual/textual. Subtitles in general add an additional level of information which viewers have to process, thereby raising the question

[2] Because of the natures of textual and verbal communication, they cannot truly communicate information identically. Most text lacks the ability to communicate specific acoustic information such as volume, timing, and other features like tone. There are some ways to represent these orthographically (such as punctuation for pauses) but strategies for this can vary (Sun & Wang, 2019). Here the word 'same' is because the text is the equivalent written form of what is spoken, so its meaning is based directly off the spoken phrase.

of what processing strategies comprehenders rely on when integrating information in two forms no matter the type of subtitle, one of the major issues with subtitling is the amount of time and resources they take to write (Orrego-Carmona, 2016). Whether it is inter- or intra- lingual, subtitles are generally considered most accurate when written by a professional. This becomes a problem because many smaller forms of media do not have the resources to be able to produce high quality subtitles. Along with this, smaller language communities are less likely to receive professionally subtitled media in their language. To solve this, there have been efforts for audiences to generate their own subtitles (as investigated by Orrego-Carmona (2016)) or have them be automatically generated by different software. Both of these options are widely used, but both are known to have many potential inaccuracies. In the case of audience generated subtitles, problems include mistakes which are not caught in editing or limited proficiency in the original language for interlingual subtitles. Auto-generated subtitles work based on speech-to-text software which detect which words are being spoken are being spoken and which depend on large amounts of data for training. These are known to have issues with any non-standard accent or vocabulary, although such problems vary by how the software was designed and trained.

Along with word recognition issues, any additional auditory information may interfere with the software, so music and sound effects can make the software's guesses less accurate. Despite these inaccuracies, the subtitles are still used because they are helpful or necessary, especially for people who need interlingual subtitles or intralingual for accessibility concerns, such as d/Deafness or processing difficulties.

Since subtitles are generally meant to aid in processing or accessibility, making them as clear as possible is a major goal. Subtitles already act as an additional source of input alongside the audio stream, so minimizing the processing power needed is a major part of making the subtitles effective. One potential issue to effective subtitle processing is when subtitles are not accurate to the dialogue. Comprehenders generally understand that subtitles should have the same information in both the auditory and textual mediums, which would not be true in the case of inaccurate subtitles. Because the information does not match their expectations, it is possible that the processing effort would be increased by the audio-textual misalignment. Considering people are still using these inaccurate subtitles, it is possible that people are developing strategies for how they view the media in order to not experience overload. In other words, they could be adapting their viewing behaviours, such as using some form of information recovery strategies, or even not reading the inaccurate subtitles.

Before these recovery strategies can be assumed and investigated, it must first be shown that inaccurate subtitles are having an effect on media processing in the first place. Currently there is very little research which has looked into how people manage conflicting information which they expect to align, so it is not clear that inaccurate subtitles impact cognitive load. This current study will investigate the effects of inaccurate subtitles on viewer's cognitive processing and viewing behaviours to see if there is a noticeable difference between contexts with accurate versus inaccurate subtitles. If inaccurate subtitles are impacting cognitive load this will then open up possible research into the behavioural changes of viewers facing misaligned information. The findings show that viewers do experience increased cognitive load in response to inaccurate subtitles, suggesting that more research is needed on viewers' behavioural responses.

2 Literature Review

The current project has been developed from a strong history of research on subtitles and cognitive processing. This section will first discuss major areas of subtitle research and how they have varied. Following from that section, cognitive load will be discussed as one of the primary factors investigated in subtitle research. Since this is often done through eye tracking studies, eye tracking as a methodology will

then be discussed, with a focus on how eye tracking is used to measure cognitive load. Because of the multi-sensory nature of subtitled content, multimodal integration will then be discussed. Finally multimodal misalignment in subtitles will be explored, as there is currently a lack of work in this area describing the effects of lexical errors in subtitle processing.

2.1 Subtitles

Over the past few decades, subtitle research has generally fallen into a few major categories: subtitles in translation, subtitles in the classroom, subtitles for accessibility, and how subtitles can be optimised for comprehension or reduced cognitive load. Each of these areas has a largely different goal in mind, with the exception of optimisation, but they tend to look at similar features such as content comprehension, attention distribution, and cognitive load. At this time no research on auxiliary subtitles has been found, though it is possible that it may have been investigated within a media studies lens.

Inter-lingual subtitles for translating foreign media are one of the most common types people encounter, given that they make media from other cultures and in other languages more widely accessible. While lots of media also undergoes dubbing, the process of replacing the dialogue with localized audio, subtitles are still often a more accessible and less labour-intensive practice (Perego et al., 2016). Inter-lingual subtitles are also occasionally created through community generation instead of by large media companies. These community practices usually involve people who are bilingual taking the time to translate programs in order to make them available where they previously were not (Orrego-Carmona, 2016). This can happen with larger media which is not professionally translated for smaller language communities or occur on smaller form content such as YouTube videos, which allows limited community subtitle generation. Research on inter-lingual subtitles spans topics related to cognitive processing (Perego et al. 2016), differences between professional and amateur subtitles (Orego et al), and bilingual subtitles (subtitles both in their original language and a translation) (Liao, Kruger, and Doherty, 2020). All of these studies analyse cognitive load and comprehension of the subtitles, with the general finding that subtitles increased comprehension and did not increase cognitive load, even in the case of bilingual subtitles. The lack of increased cognitive load from bilingual subtitles is interesting because it suggests that the additional information stream from the subtitles does not necessarily interfere with other elements of processing, suggesting it may be due to viewers prioritising the language they are familiar with over the auditory information which is unfamiliar (Liao, Kruger, and Doherty, 2020). This pattern was noted particularly in inter-lingual subtitles in classroom settings, where the primary language of instruction was found to be prioritised even over students' native language (Kruger, 2013).

Subtitles in classroom settings are more often intra-lingual and the research on them usually has a strong focus on content comprehension and attention. Generally, the studies look at the usefulness of subtitles as a tool which can be added to video recordings and virtual lectures, though some research has also been done on how pre-existing subtitled media can be used for increasing student literacy (Goldman & Goldman, 1988). When looking at comprehension, multiple studies have found a significant increase in comprehension rates when subtitles are employed (Kalyuga, 2012; Kruger, 2013; Yoon & Kim, 2011). This is usually measured by comparing test scores of students who were exposed to subtitled classroom content and those who were not. In terms of how attention is allocated on screens, it has been found that subtitles lead to an increase in students' attention to non-subtitle areas which likely also contributes to the increase in comprehension. Classroom research has also been a primary area where subtitle optimisation has been relevant, specifically with the aim of helping students retain as much information as possible. Research in this area does tend to make the assumption that subtitles are accurate, which leaves open a question of how these findings are impacted by subtitle inaccuracies.

Subtitles as a tool for accessibility are an interesting area because they exist between classroom use and translation. Most of the research in this vein focuses on Deaf and hard of hearing individuals because of the issue presented by largely auditory dialogue. For many Deaf individuals their first language is a Sign language, so they may have lower proficiency in the spoken language of their larger community, even in its written form. This means not only that subtitles need to be present to make media accessible, but also issues of clarity and style for optimised comprehension are pertinent. Research in this area thus has similar focuses on comprehension and attention like the previous two categories (Yoon & Kim, 2011) but also examines what styles of subtitles are most useful. For instance, Szarkowska et al. (2011) looked at how verbatim, edited, or standard subtitle types affected hearing, hard of hearing, and Deaf viewers differently. They found that overall, all groups had higher dwell time on verbatim subtitles, with Deaf participants spending more time than the other two groups on verbatim subtitles. Because higher dwell time is linked with increased cognitive, this means Deaf participants' processing is more impacted by verbatim subtitles than their hearing and hard of hearing counterparts. The significant difference of verbatim subtitles is likely because verbatim subtitles contain disfluencies and repairs which would make them less grammatical when written. These findings suggest that the less grammatical form of verbatim subtitles does increase the processing effort for people less accustomed to the grammatical errors of spoken speech, such as d/Deaf individuals. These findings show that less grammatical styles of subtitles (in this case verbatim subtitle) can have a direct impact on cognitive load, which may apply to other styles. In the current research lexical errors found in autogenerated subtitles are being examined for their impact on cognitive load. The focus on how subtitles are styled leads directly into the final major area of research, roughly categorised here as subtitle optimisation.

Given that certain subtitle styles are known to cause reading disruption, an open question is how best to optimise subtitles. Most of the research on the production and design of optimised subtitles focuses on reducing cognitive load and achieving high comprehension rates. Factors that have been investigated about optimised subtitle design include things such as the number of lines per subtitle (Szarkowska & Gerber-Morón, 2019), the subtitle speed (Kruger, Wisniewska, & Liao, 2022), or the style of subtitles employed (here verbatim, edited, and standard subtitles) (Szarkowska et al., 2011). This research has generally found that two-line subtitles which match the timing of the dialogue produce the highest level of comprehension according to post-experiment comprehension test results. In terms of subtitling "styles," one example of a style type prominent in the research is integrated subtitles. Integrated subtitles are a style of subtitle where the text is not placed in the bottom third of the screen. They are instead placed closer to the visual focal points in the video (Kruger et al., 2018). Usually this involves moving the dialogue closer to its' speaker and making sure the text does not block anything relevant in the image. Along with this, some integrated subtitles will use subtitle colour and opacity to signal volume or tone. Overall, research on integrated subtitles found that they potentially reduced cognitive load as measured via a decrease in fixation count compared to conventional subtitles (Kruger et al., 2018; Black, 2022). These are some examples of how subtitles can be optimised in how they are designed and presented to viewers, especially in relation to cognitive load management.

2.2 Cognitive Load Theory

Originated by John Sweller, cognitive load theory posits that people's working memory is a limited resource, and this limit influences the amount of information which can be processed at a time (Sweller, 1988). Cognitive load is generally framed as different stimuli's properties requiring different amounts of cognitive effort to process them. To explain these properties, cognitive load is usually divided into three types (Kruger et al., 2018):

1. Extraneous load: The processing load which comes from how information is presented. Extraneous load is composed of elements which increase processing effort but not necessarily comprehension or information retention. Because it is related to presentation, the amount of extraneous load can be manipulated. An example of manipulating extraneous load would be the number of lines subtitles are presented on, as mentioned above (Szarkowska & Gerber-Morón, 2019). Three-line presentation had an increased extraneous load, which resulted in lower comprehension rates.
2. Intrinsic load: The processing load associated with the difficulty of information in reference to the prior knowledge of the learner/experiencer. A variety of factors influence the amount of intrinsic load, but they cannot be manipulated directly. For instance, if a person is a trained artist, it will take much less effort for them to learn a new artistic medium than someone completely unexperienced.
3. Germane load: The processing load and mental resource allocation for consolidating information into mental schema, associated with overall processing. Mental schemas are thought patterns which organise information based on categories and how they relate to each other. Germane load is the description of the cognitive processes which go into forming these mental schemas. This type of load can be manipulated, primarily through reducing the amount of extraneous load being processed.

While these types of cognitive load are important for understanding how different stimuli impact processing effort, it is important to remember that measures of cognitive load are not able to fully account for individual differences. An example of this is measuring how much processing goes into schema formation because it is largely based on participants' existing knowledge (Doherty & Kruger, 2018). Despite the difficulties individual differences present, it is one of the main theories used to understand how people process stimuli which involve multiple streams of input simultaneously. Cognitive load is usually studied in reference to educational strategies, though it has expanded into other areas of psychological research as well. This is because of the relevance of learning information and seeing how best to present it for the information to be retained. Because cognitive load is entirely internal, it is currently not possible to measure directly. Instead, external realizations are investigated to assess it comparatively, such as comparing eye movement patterns in response to more and less complex stimuli.

2.3 Eye Tracking

One of the main research methods used to analyse cognitive attention is eye tracking. Eye tracking is the process of measuring where people look in a variety of environments/stimulus conditions. One major way this is analysed is using pupil contrast and corneal reflections to track movement, though electric potential of eye-muscles or eye-attached equipment is also used. Eye tracking is used for a variety of research types such as reading studies (Cook & Wei, 2017), reaction time analysis (van Ens et al., 2019), or marketing research (Wedel & Pieters, 2017).

This stems from the strong eye-mind hypothesis which proposes that gaze fixations reflect where people are focusing their attention (Underwood & Everatt, 1992). Eye tracking generally functions by tracking fixations, which are static dwell points, along with saccades, the eye movements between fixation points. Along with these measures, some equipment is able to measure pupillary contractions or retinal vein tracking, though these are less common (Doherty & Kruger, 2018). Measuring attention through these factors has opened up interesting methodological approaches in multiple disciplines.

One major area within psycholinguistics where eye tracking has been used is in reading studies, which look at eye movement patterns when participants are reading text. Along with generally understanding how gaze movement works during reading, this has been used to investigate a variety of different features such as textual processing speed (Gwizdka, 2014) or how people parse new vocabulary (Godfroid et al., 2017). Even though studies on subtitles are related to reading studies because of the focus on processing textual information, there is competing sensory information not present in most reading studies. Along with this, the addition of other visual and auditory stimuli makes linking them directly to cognitive attention more difficult. Studies on subtitle processing have been increasing in prevalence in recent years, both due to an increase in subtitle availability and eye tracking equipment becoming more efficient and accessible. The latter has been due to increases in available eye tracking technologies, whether it is cheaper desktop setups or software which employ computer webcams (Kok et al., 2023). When looking at subtitles, there are multiple reasons for using eye tracking such as investigating attention, motivation, and cognitive load. Here we will be focusing on cognitive load in regard to the multi-sensory input of subtitled media.

2.4 Measuring Cognitive Load in Eye Tracking

Because of the strong eye-mind hypothesis, eye tracking is able to directly investigate cognitive load responses. One of the primary measures which is linked to cognitive load is pupillary response, with dilation being associated with high cognitive load (Gwizdka, 2014). This is due to decreases in parasympathetic activity in the brain, which impact dilation rates. Because pupillary responses to light level are hard to distinguish from other varieties of responses, this measure is limited to specific stimuli. The variance in light in media such as movies or television shows can interfere with pupillary tracking, even when environmental light levels are consistent. Because of this, subtitle research does not generally employ pupillary response as a factor.

Two other factors often investigated are fixation count and dwell time (also called fixation duration). Higher fixation counts are associated with heightened cognitive load because they imply more information being processed (Kruger, Szarkowska, & Krejtz, 2015). Longer dwell times are also associated with higher cognitive loads, with Holmqvist et al. finding that longer fixations correlate with ‘deeper’ processing (2011). In the literature deeper processing was not properly defined, but correlates with higher levels of cognitive load, which is how it will be approached here. This more intensive processing type is seen as a way of tracking increases in general cognitive load. Here the factors fixation count and dwell time will be categorized as cognitive measures of study. Some other common factors analysed in eye tracking are based off these, such as area revisits and visit %. Here these will be referred to as behavioural measures because they do not directly track the cognitive processes of viewers, but do measure how they interact with the media they are consuming.

Another way in which attention is analysed in reading studies is saccades/gaze paths, which look at what information is prioritised and in what order (Gwizdka, 2014). Because subtitles are not static, traditional reading study methods which measure saccades between single-word fixations become harder. The difficulty is both due to additional visual input and because viewers have to prioritise certain word types in the limited timeframe (Kruger, Szarkowska, & Krejtz, 2015). In order to account for the dynamic nature of subtitles, Kruger and Steyn developed the Reading Index for Dynamic Texts (RIDT) (2014). This index uses reading study measures of comprehension, including saccades patterns across words, while accounting for the additional visual focus points in subtitled media. The RIDT is calculated using the number of unique fixations on standard words compared with the average saccade length on that subtitle based on average word length. This index is just one example of the ways research has been adapting to better understand the complex input associated with subtitled media.

The current research will be prioritising fixation count and dwell time as measures of cognitive load because comprehension is not a goal of the current research project. Since they are measures of cognitive load, they are also more responsive to the multimodal nature of subtitled content.

2.5 Multimodal Integration

Multimodal integration is the term for a combination of cognitive processes which simultaneously combine information from separate sensory sources for the brain to process (Kruger et al., 2018). With multiple streams of input, the brain must use collections of inferences and existing schematic connections in order to process. An example of this would be the sensory input received from fireworks. Fireworks produce a brief concussive sound along with bright lights which may be in different shapes or colours. While the concussive sound could be confused with other sources, such as thunder or firearms, the addition of the visual input narrows down the source. Thunder may have an associated bright light in the form of lightning, but this would be a distinctive shape and short duration, while firearms do not have associated light. By processing the multiple sensory input sources using existing schema, comprehenders are able to quickly understand that fireworks are happening near them. In terms of cognitive load theory, multiple sources of input increase the amount of cognitive load, which means people must either miss some of the information or try to adapt their processing behaviours to manage the increased input without becoming overloaded (Kruger et al., 2018).

Within multimodal theories, one of the most relevant components to subtitle processing is the verbal redundancy effect. In multimodality, redundancy is when two input sources are delivery the same information (Kalyuga, 2012). While this often leads to increased cognitive load, this does not seem to hold for verbal redundancy. Verbal redundancy, identical linguistic information produced as multiple separate inputs, is usually discussed in terms of simultaneous signing and speaking. However, another major form is found in subtitles and other read-along texts, which present the information as both written and spoken input. In both of these instances, verbal redundancy has been found to have no increase on cognitive load in comparison to a singular input (Kalyuga, 2012). In fact it has been linked with higher levels of content comprehension, suggesting that the two input streams may in fact reinforce each other, which would aid in schema creation (Kruger, Szarkowska, & Krejtz, 2015). It is believed that this occurs because the multiple streams of linguistic information are in fact the same despite their different forms, which then leads to questions of what happens when it does not align perfectly, such as inaccurate subtitles.

2.6 Multimodal Misalignment

In current research, inaccurate subtitles as a phenomenon are largely unexplored. Some research however has looked at how different subtitle styles impact comprehension and cognitive effort. One such study was conducted by Perego et al. (2010) which tested phrase segmentation in Italian subtitles of a Hungarian television show. They investigated whether viewers responded differently to the separation of an adjective onto a different line from its associated noun. In the end, such adjective-noun separation was not found to have any effect. Other investigations have looked at style of subtitles, such as Orrego-Carmona (2016) and Szarkowska et al. (2011). In the former, professional and amateur translations were compared and no processing difference was found between the two. The latter looked at verbatim, edited, and standard subtitles, with the findings being explained above. While these subtitle conditions do use subtitles whose content did not fully match what was said or which used a non-standard translation, neither study actually addresses the issue of inaccuracies directly. The lack of research in this area is especially salient with autogenerated subtitles, which are increasingly common and often have lexical inaccuracies. Here lexical

inaccuracies are defined as words found in the subtitles which do not have the form as what was spoken in the dialogue. Usually, these inaccuracies occur because of similar phonology being processed by the auto-generation software, leading to the wrong words being output as subtitles. Thus far, these lexical inaccuracies and their impact on viewers' processing of media has been unexplored, so they will be investigated here.

3 Method

3.1 Participants

An availability sample of 29 participants was collected with an average age of 21 (SD = 1.374). One participant had to be excluded due to a strong glasses prescription interfering with the eye tracker's ability to find their corneal reflection, so only 28 participants were included in the final data. Participants were collected through convenience sampling using the platform Instagram. The study was advertised as an eye tracking study on subtitles, which would accept any fluent English speakers. The choice to not only investigate native English speakers was made because there is a higher chance non L1 consume media using subtitles, which can't be guaranteed of English L1 speakers in an English-speaking country. The aim of this was to have a participant group better representative of the average subtitle user, thus including both L1 and L2 English speakers. All the participants were students at the University of Edinburgh, Scotland.

64.3% of participants were Native English speakers. Through a classification based on the Common European Framework of Reference for Languages (CEFR) 89.3% of the participants self-reported Advanced-High Proficiency and the remaining 10.7% reported Advanced. Participants were also asked to self-report on if they had any forms of physical or cognitive processing factors which may affect their comprehension of sensory information. While 0% of the participants had any physical factors, 39.3% reported some form of cognitive factors. Due to the limited scope of this study and out of respect for participants' medical privacy, specific details of these factors were not collected. No participants were excluded from any aspect of analysis due to any cognitive processing issues because they represent a large population group known to use intralingual subtitles (Simpson, Dalal, & Senaan, 2022). For the eye-tracking component, all participants reported normal or corrected-to-normal vision. As a part of the post-experiment questionnaire, they were also asked about their usual subtitled media consumption behaviour, but due to confusion expressed by the participants this data will not be considered here. As compensation for their time, participants were offered baked goods provided by the researcher.

3.2 Materials

For the video stimuli, a 10-minute section of the Nat Geo Wild nature documentary Wild Untamed Brazil was selected (see Figure 1). The documentary presents different animal populations which live in the Mata Atlântica forest in coastal Brazil, focusing on the animals' behaviours and the habitats they inhabit. This documentary was selected because of its clear narration, which had a standard RP accent which would be familiar to students in the UK. Along with auditory features familiar to the participants, the entire video is narrated, so no individuals are visibly speaking. This was an important factor so participants would not be able to use any misalignment recovery strategies, such as using lip reading, facial expressions, or body language (for further discussion on using narrated content, see Kruger, Szarkowska, & Krejtz (2015)).

For the experiment, each participant saw one version of the video: the original un-subtitled video, a version with YouTube auto-generated subtitles (containing lexical errors), and a version with human-corrected subtitles (no errors).

Figure 1: *Original Unmodified Video File: Documentary with no Added Subtitles (Wild Untamed Brazil, 00:00:06).*

For the un-subtitled version, the original video file was left unedited. For the stimulus with subtitle inaccuracies, the auto-generated subtitles were manually corrected so that only lexical errors were present, with a total of 21 lexical errors. For the stimulus with only accurate subtitles, the auto-generated subtitles were manually corrected, making them accurate to the narration. Due to human error, three lexical errors were still present in the third stimulus at the time of the experiment. This was taken into account in the final results, detailed in the analysis plan below.

All subtitles in both stimuli were coded into .srt files, which were then integrated directly into the video files. Due to an error in the file transfer, subtitle #43 was not visible in either stimulus, but since this occurred in both subtitle conditions it did not impact the data analysis plan. The subtitles were integrated into the video using the online service VEED (https://www.veed.io/).

Due to cost restrictions, a watermark was visible in the top-left corner of the autogenerated subtitles and the corrected subtitles, but visual inspection confirmed that this was not a common interest area for participants over the 10-minute duration. For maximum visibility throughout the entire video, the subtitles were presented as sans-serif, bold white text. They were also backgrounded by a transparent grey backing for higher visual contrast. The subtitles were placed in the lower third of the screen, so they did not interfere with the focus points of the video and so the subtitle position aligned with the standard subtitles participants are accustomed to (as opposed to another style such as integrated subtitles). Samples of the accurate and inaccurate subtitle versions are shown in Figures 2 and 3.

Figure 2: *Corrected Subtitles: Documentary with Integrated Human-Corrected Subtitles (Wild Untamed Brazil, 00:00:06).*

Figure 3: *Autogenerated subtitles: Documentary with Lexical Errors Present in the Subtitles (Wild Untamed Brazil, 00:00:06).*

All three stimuli were uploaded into Experiment Builder for presentation to participants during the experiment.

3.3 Procedure

Before being invited to the lab, a copy of the experiment information and consent form were provided to participants to review. After signing the consent forms, the participants were given an anonymized participant number and distributed into one of the three stimulus conditions: The first group (n = 9) was assigned to the stimulus without subtitles, the second (n =9) watched it with embedded autogenerated subtitles including lexical errors, and the third (n =10) watched it with human-corrected subtitles. Before watching the video, participants were informed that they would be completing a demographic and comprehension survey after the video finished. The survey used was a Google Forms questionnaire which

was separated into two components. The first component was a series of demographic questions, the results of which are detailed in the Participants section above. The second component was a series of comprehension questions asking specific content from the video watched during the experiment. These questions related to very specific details from the dialogue of the stimulus, such as "What colour are female Blue Manakins?"

The data collection began with participants being briefed on the eye tracking equipment and stages of the experiment. An important aspect of this was making participants aware that they would be asked comprehension questions after the recording period. Making them aware of the comprehension survey was specifically emphasized to encourage participants to focus on video's content, which would provide more natural data than if they were consciously primed that the data collection would be about their subtitle attention. The eye tracking system used was an Eyelink 1000 Plus in Tower Mount configuration, using the headmount. The monitor used had a resolution of 1980 x 1020 and was placed 60 centimetres away from the eye tracker. Monocular data was collected from participants' right eyes. All recording settings were set to the system default and the experiment was designed for the monitor specifications. Throughout all trials a consistent light level was maintained in the room.

The first stage was to calibrate the eye tracker, which was individually done for each participant before the recording began. A 9-point device calibration phase was integrated into the experiment along with a secondary verification stage. Once the calibration was completed, the experimenter began the recording phase. In the recording phase, participants were shown one of the three stimulus videos as the eye tracker recorded their gaze. This data was then recorded as an .edf file. The final phase of the experiment was having participants complete the Google Form which asked them demographic and comprehension questions.

3.4 Analysis Plan

For data analysis, the first thing done was uploading the experiment data files into Data Viewer. Here they were sorted and aggregated by stimulus video. Areas of interest (AOIs) were created for each line of the subtitles, as shown in Figure 4.

Figure 4: *Autogenerated Subtitle Video Interest Areas, Shown with Fixation Points.*

Initially AOIs were also going to be created on the primary visual focus points in each section of the video, but due to time constraints this was deemed beyond the scope of the current project. Because visual focus points were not analysed, data from the un-subtitled group was not used in the final analysis. Additionally,

the subtitles which were affected by errors in the design stage were not included in the analysis in either condition. The data analysed from the interest areas was fixation number, dwell times, trial visit %, and trial runs. Fixation number is the average number of individual fixations which fall inside the AOI. Dwell time is the average time spent on each fixation within the AOI. Trial visit % is the number of participants who had at least one fixation within the AOI, separated by condition. Trial runs, also called revisits, are instances where a fixation occurs inside an AOI followed by one outside and then another inside. These imply that a viewer is focusing on something outside the AOI but returning, which tracks split attention. These are the same measures used in the analysis done by Orrego-Carmona on professional versus non-professional subtitling, as mentioned above (2016).

The analysis targets participants' looking behaviour in the 21 video scenes whose subtitles differ across the accurate and inaccurate versions of the video. The analysis uses averages computed for each scene (computed by averaging across participants), for each of the following four measures: dwell time, fixation number, trial visits, and trial runs. For each measure, a 2-tail paired T-Test was conducted to compare the accurate and inaccurate conditions. In addition, linear regression models were created to examine how participants' behaviour changed throughout the video, using subtitle number as an approximate measure of time. The time analysis was done due to a trend noticed in the data, which prompted further investigation, and provides an opportunity to see whether participants who are exposed to a video with inaccurate subtitles adopt different viewing strategies as the video progresses.

4 Results

Table 1: *Dependent variable statistics*

	Dwell Time (ms)		Fixation Count		Run Count		Visited Trial %	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Corrected	1459.427	810.562	7.332	4.028	2.762	1.34	87.5	9.798
Incorrect	1706.819	911.878	7.447	3.934	2.607	1.4	79.856	11.339

For each of the four dependent variable variables (dwell time, fixation count, run count, and trial visit percent) a 2-Tailed Paired *T*-Test was conducted using the averages of the corrected and inaccurate conditions across the 66 subtitles in the stimuli. All mean and standard deviation values are listed in Table 1 above.

For Dwell Time, inaccurate subtitles yielded longer dwell times than corrected subtitles ($t(131) = 4.0691$, $p = .0001$). Fixation Count was found to not be significant ($t(131) = 0.5440$, $p = 0.588326123$). For Run Count, corrected subtitles produced more average runs than inaccurate subtitles ($t(131) = 2.0092$, $p = 0.0487$). For Visited Trial % inaccurate subtitles yielded lower viewer percentages than corrected subtitles ($t(131) = 5.2548$, $p = 1.7591E-06$).

Additionally, for all of the dependent variables a linear regression was created with fixed effects for subtitle number (an index of time, included in the model as scaled subtitle number), condition (coded with "corrected" set to -.5 and "incorrect" set to .5) and their interaction. No random effects were included.

The analysis shows that dwell time is lower for correct subtitles and decreases over time (over subsequent subtitle numbers). Both of these main effects are statistically significant (time: B=-12.254, SE= 3.529, t=-3.472, p<0.001; condition: B= 627.127, SE= 284.482, t= 2.204, p<0.05). There is no interaction (B=-10.911, SE= 7.058, t=-1.546, p=0.12).

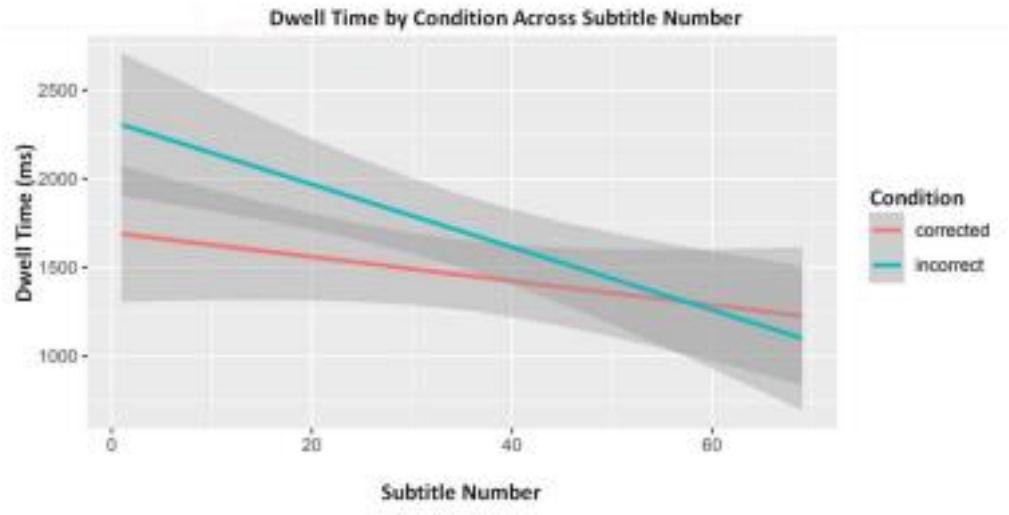


Figure 5: *Dwell time by condition using subtitle number as an approximation of time.*

For fixation count the counts decrease significantly over time (*B*=-0.04275, *SE*= 0.01671, *t*=-2.558, *p*<0.05). There is no significant effect of condition nor an interaction (*t* < 1).

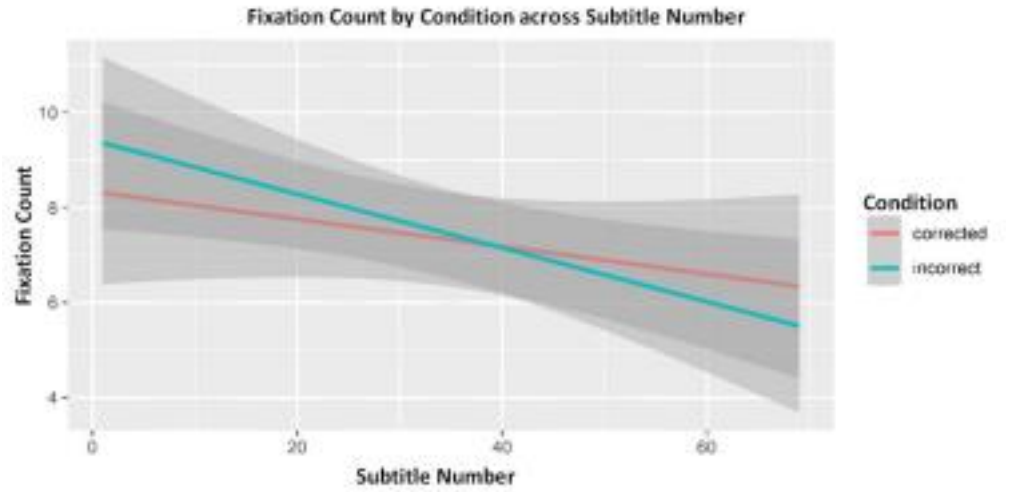


Figure 6: *Fixation count by condition using subtitle number as an approximation of time.*

For run count the counts decrease marginally over time (*B*=-0.009864, *SE*= 0.005846, *t*=- 1.687, *p*=0.09). There is no significant effect of condition nor an interaction (*t* < 1).

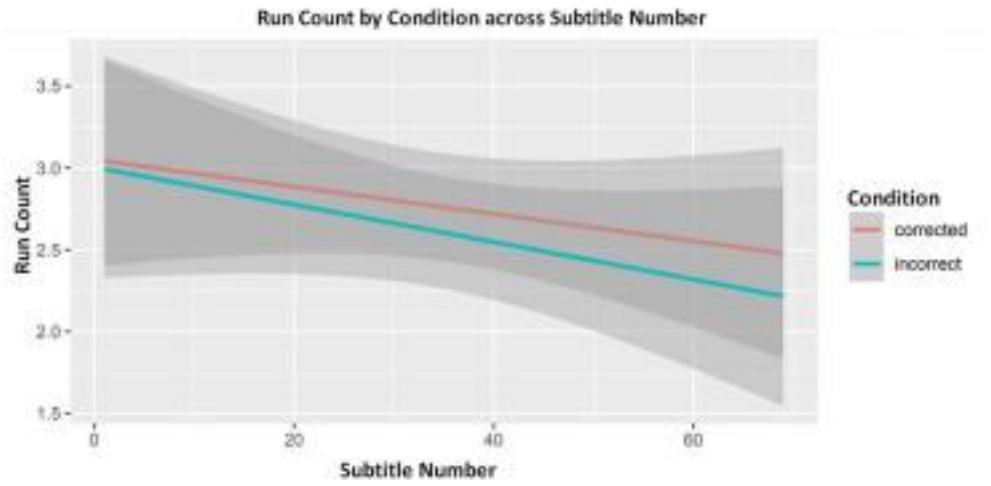


Figure 7: *Run count by condition using subtitle number as an approximation of time.*

For trial visit % the counts decrease significantly over time (*B*=-0.09389, *SE*= 0.04444, *t*=-2.113, *p*<0.05) an effect driven by a marginal interaction (*B*=-0.15618, *SE*= 0.08888, *t*=- 1.757, *p*=0.08), whereby the decrease is more apparent for incorrect subtitles than correct subtitles. There is no significant effect of condition ($t < 1$).

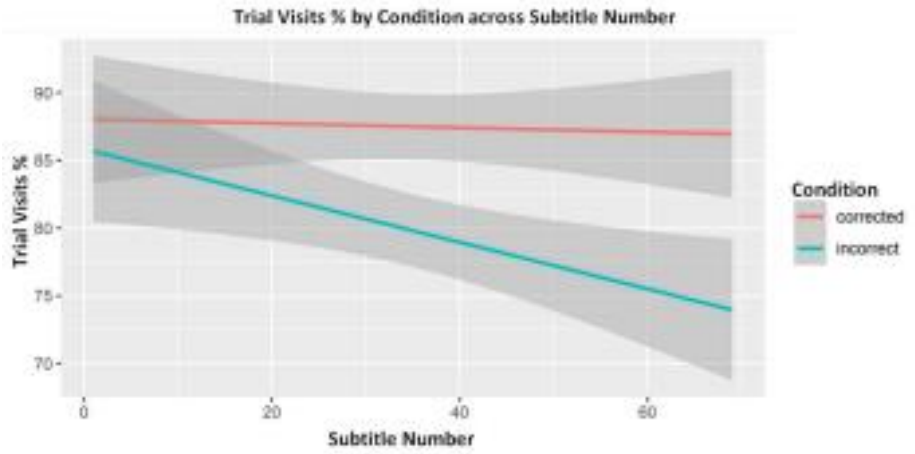


Figure 8: *Trial visits % by condition using subtitle number as an approximation of time.*

5 Discussion

This experiment aims to explain the cognitive effects of lexical errors in subtitles. By using eye tracking measures, it is possible to look at how multimodal misalignment from verbal stimuli impacts cognitive load. Here we have found that cognitive load is being impacted by inaccurate subtitle conditions, shown through an increase in dwell time. Along with this, participants also alter their viewing behaviours in response to the subtitle accuracy. Additionally, all of the factors investigated showed an effect of time which requires further analysis.

In our discussion of this, we will first discuss each of the eye tracking measures investigated and how these results compare with similar research. The eye tracking measures relating to viewer behaviours will be discussed separately from measures which are theoretically linked directly to cognitive load. We will then discuss the implications of the linear regression models and the need for further analysis of behavioural changes over time in future research.

5.1 Eye Tracking Measures: Fixation Count

In the results, fixation counts were not found to vary by conditions. This is interesting because it differs from the investigation of integrated subtitles by Kruger et al. (2018). Because their research is also on the cognitive load associated with intralingual subtitles, it was expected that similar results would be found. In their research, fixation count was found to be one of the primary significant eye tracking measures, which they linked to extraneous load. What is not clarified in their study is how fixation counts are connected to extraneous load specifically, while dwell time and other factors are measures of average cognitive load at the same time.

Following this idea, the fact that fixation counts are not significant would suggest that additional extraneous load is not impacting the overall cognitive load as would be expected. In this experiment, fixation count did not differ significantly between the conditions, but the differences in other measures suggest that cognitive load is still being impacted. The nonsignificant effect of fixation duration across conditions would imply that lexical errors do not impact the extraneous load, which is interesting because it is the primary form of cognitive load which is known to be manipulated. Some possible reasons for this will be discussed below alongside the other measures.

5.2 Eye Tracking Measures: Dwell Time

Dwell time is taken to indicate average cognitive load and ‘depth’ of processing (Kruger, Szarkowska, & Krejtz, 2015). Even though processing depth is generally an unclear concept within cognitive load theory, it is clear that it has connections to higher overall cognitive load, but not any specific form of it. Generally, the term ‘depth’ seems to suggest that more processing effort is allocated with the assumption that more comprehension would be gained from this, but this is currently not directly stated in any research found. Here dwell time will be discussed in terms of general cognitive load and the potential connections to either intrinsic or germane load.

In the data, dwell time was found to be significantly longer in the inaccurate condition, suggesting that lexical errors do indeed impact cognitive load for viewers. Interestingly, in the same Kruger et al. study mentioned above, dwell time showed no differences by cognitive load even though fixation count did (2018). Because Kruger et al. was investigating integrated subtitles when they found dwell time to have no effect, it is hard to say whether this means that the increased cognitive load is specifically linked with lexical errors, because dwell time has been shown significant in other studies like Szarkowska et al. (2011). Their study included subtitling styles which did not directly match the audio, so their results show that the increase has been found in similar environments elsewhere. The fact that dwell time was significant while fixation count was not could be linked to the fact that fixation count is usually specifically linked with extraneous load, which is a result of the presentation of stimulus information. It also cannot be due to intrinsic load because theoretically it cannot be manipulated because of its relation to the actual informational content of the stimulus. Since it is neither of the other forms of cognitive load, lexical errors may instead be interfering with germane load, which deals with schema formation processes. One possible reason for the impact on germane load could be because lexical errors require viewers to infer what words

or concepts were meant to be written, or to process only the auditory input. At this current stage this is purely speculation and further research will be needed to try and understand this connection.

5.3 Eye Tracking Measures: Visit % and Runs

These two measures are being reported together because neither are typically theoretically linked measures of cognitive load. Instead, they are measures which show the behaviour patterns of viewers under the different conditions. The data showed that both measures differed significantly between the accurate versus inaccurate conditions, with the presence of lexical errors resulting in lower run counts and fewer visits across participants. This suggests that participants are in fact behaving differently under different subtitle conditions. In the case of visit % this means that fewer participants were even reading the subtitles with errors, which suggests that they may not read 'unreliable' subtitles as often. In the case of runs, it is hard to determine because the run count would also be responsive to trial visit %, though it definitely appears that there is a behavioural change. This will be further discussed in the discussion of the behavioural changes over time.

5.4 Viewing Behaviour over Time: Cognitive Measures

The linear regression models are interesting to investigate because time connected analyses are almost unexplored across the literature. Perego et al. (2010) did an analysis of behaviour changes between the first half of a video and the second, but no other studies were found to have any measures of changes across time. Perego et al. reported that in the second half of the video participants increased their fixation counts and dwell time in non-subtitled regions. This does not fully align with the findings here, but because it was an inter-lingual subtitle study it is not surprising that participants responded differently.

In this data, both of the cognitive measures were found to decrease significantly over time. The fact that fixation count decreased over time even though fixation counts didn't differ significantly by subtitle accuracy in the overall data suggests that viewers' processing naturally varies the longer they are exposed to a stimulus. This shows that even under normal subtitled conditions there appears to be a change occurring the longer viewers watch a stimulus. Along with this, dwell time also varied which suggests that the overall cognitive load was varying across time as well. These findings show that more research needs to be done to establish a baseline of how these cognitive measures vary across time.

5.5 Viewing Behaviour over Time: Behavioural Measures

In the analysis of response to subtitle accuracy over time, the behavioural measures also showed a significant change. Additionally, the visit % showed marginal interaction in which the decrease for incorrect subtitles was more apparent than for the corrected condition. This suggests that the lexical errors are causing a difference which is separate from the effect of time alone.

Because this is a measure of how many participants viewed the subtitles, this decrease over time suggests that viewers of inaccurate subtitles are choosing to not process the competing information. If this is true, the misaligned input could be causing people to prioritise the auditory information because it is more accurate. This requires further examination because it is hard to determine how this the impacts the accuracy of the other factors.

Because all of the measures were taken from fixations within the AOI, if participants are not reading subtitles as often over time, then the measures from later sections of the video are not accurate. Because of the limited scope of this research, it was not investigated if the same participants tended not to

look at inaccurate subtitles or if it is just the average across participants. Future research needs to be done with an awareness of participants who are not looking at the subtitling region, especially if it is a behaviour chosen by specific individuals. This could be done by an analysis which also looks at these measures on the visual focus points of the videos along with the subtitles. This would show if the cognitive load factors are consistent when prioritising visual, non-textual input.

6 Conclusion

While accurate subtitles are assumed in all subtitles, manual and autogenerated, this is not the reality experienced by viewers. The prevalence of inaccurate subtitles is why it is important that research is not only conscious of this, but actively investigating the effects. Here we have done the initial work by showing that lexical errors both increase cognitive load for viewers and change their viewing behaviour patterns. Establishing a baseline of how viewers respond to lexical errors is a key first step to understanding how people cope with misalignment in co-occurring stimuli. We have also shown that lexical errors do impact viewing behaviours in response to cognitive load as well. While here the data suggests that people may abandon inaccurate stimulus sources as a strategy to cope with misaligned input, this needs to be further investigated. Along with this it is important to see how people utilize other information streams like body language, lip reading, and visual context to recover information. Understanding how multimodal input is integrated and processed will involve further experiments, which could use obstructing different video and audio components to see how different conditions affect cognitive load. This current project has been the first step in understanding how multimodal misalignment affects viewers, which will hopefully help motivate accurate subtitle sources and more awareness of how viewers process information in media.

7 References

Black, S. (2022). Could integrated subtitles benefit young viewers? Children's reception of standard and integrated subtitles: a mixed methods approach using eye tracking. *Perspectives: Studies in Translation Theory and Practice*, 30(3), 503–519.

Brahnam, S. (2017). Comparison of in-person and screen-based analysis using communication models: A first step toward the psychoanalysis of telecommunications and its noise. *Psychoanalytic Perspectives*, 14(2), 138-158.

Cook, A. E., & Wei, W. (2017). Using eye movements to study reading processes: Methodological considerations. In C. Was, F. Sansosti, & B. Morris (Eds), *Eye-tracking technology applications in educational research* (pp. 27-47). Hershey: IGI Global.

Doherty, S., & Kruger, J. L. (2018). The development of eye tracking in empirical research on subtitling and captioning. In T. Dwyer, C. Perkins, S. Redmond & J. Sita (Eds), *Seeing into screens: Eye tracking and the moving image* (pp. 46-64). London: Bloomsbury.

van Ens, W., Schmidt, U., Campbell, I. C., Roefs, A., & Werthmann, J. (2019). Test-retest reliability of attention bias for food: Robust eye-tracking and reaction time indices. *Appetite*, 136, 86-92.

Godfroid, A., Ahn, J., Choi, I. N. A., Ballard, L., Cui, Y., Johnston, S., ... & Yoon, H. J. (2018). Incidental vocabulary learning in a natural reading context: An eye-tracking study. *Bilingualism: Language and Cognition*, 21(3), 563-584.

Goldman, M., & Goldman, S. (1988). Reading with close-captioned TV. *Journal of Reading*, 31(5), 458-461.

Gwizdka, J. (2014, August). Characterizing relevance with eye-tracking measures. *Proceedings of the 5th information interaction in context symposium*, 58-67.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford: Oxford University Press.

Kalyuga, S. (2012). Instructional benefits of spoken words: A review of cognitive load factors. *Educational Research Review*, 7(2), 145-159.

Kok, E. M., Jarodzka, H., Sibbald, M., & van Gog, T. (2023). Did You Get That? Predicting Learners' Comprehension of a Video Lecture from Visualizations of Their Gaze Data. *Cognitive Science*, 47(2), e13247.

Kruger, J. (2013). Subtitles in the classroom: Balancing the benefits of dual coding with the cost of increased cognitive load. *Journal for Language Teaching*, 47(1), 29-53.

Kruger, J. L., Doherty, S., Fox, W., & de Lissa, P. (2018). Multimodal measurement of cognitive load during subtitle processing: Same-language subtitles for foreign-language viewers. *Innovation and expansion in translation process research*, 18, 267-294.

Kruger, J. L., & Steyn, F. (2014). Subtitles and eye tracking: Reading and performance. *Reading Research Quarterly*, 49(1), 105–120.

Kruger, J. L., Szarkowska, A., & Krejtz, I. (2015). Subtitles on the moving image: An overview of eye tracking studies. *Refractory: a journal of entertainment media*, 25, 1-14.

Kruger, J. L., Wisniewska, N., & Liao, S. (2022). Why subtitle speed matters: Evidence from word skipping and rereading. *Applied Psycholinguistics*, 43(1), 211–236.

Liao, S., Kruger, J. L., & Doherty, S. (2020). The impact of monolingual and bilingual subtitles on visual attention, cognitive load, and comprehension. *The Journal of Specialised Translation*, 33, 70-98.

Nat Geo Wild. (2014). *Wild Untamed Brazil* [Video]. https://www.natgeotv.com/za/shows/nationalgeographicwild/wild-untamed-brazil.

Orrego-Carmona, D. (2016). A reception study on non-professional subtitling: Do audiences notice any difference? *Across Languages and Cultures*, 17(2), 163–181.

Perego, E., del Missier, F., Porta, M., & Mosconi, M. (2010). The cognitive effectiveness of subtitle processing. *Media Psychology*, 13(3), 243–272.

Simpson, E., Dalal, S., & Semaan, B. (2022). " Hey, Can You Add Captions?": The Critical Infrastructuring Practices of Neurodiverse People on TikTok. *arXiv preprint* arXiv:2212.06204, 1-27.

Sun, K., & Wang, R. (2019). Frequency distributions of punctuation marks in English: Evidence from large-scale corpora. *English Today*, 35(4), 23-35.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257-285.

Szarkowska, A., & Gerber-Morón, O. (2019). Two or three lines: a mixed-methods study on subtitle processing and preferences. *Perspectives: Studies in Translation Theory and Practice*, 27(1), 144–164.

Szarkowska, A., Krejtz, I., Klyszejko, Z., & Wieczorek, A. (2011). Verbatim, standard, or edited? reading patterns of different captioning styles among deaf, hard of hearing, and hearing viewers. *American Annals of the Deaf*, 156(4).

Underwood, G., & Everatt, J. (1992). The role of eye movements in reading: Some limitations of the eye-mind assumption. *Advances in psychology*, 88, 111-169.

Veed - Online. VEED.IO. (n.d.). Retrieved February 25, 2023, from https://www.veed.io/.

Wedel, M., & Pieters, R. (2017). A review of eye-tracking research in marketing. *Review of marketing research*, 123-147.

Yoon, J. O., & Kim, M. (2011). The effects of captions on deaf students' content comprehension, cognitive load, and motivation in online learning. *American Annals of the Deaf*, 156(3), 283-289.

8 Acknowledgments

This research has been such a big part of my life this past year, and it would not be what it was without all the wonderful people in my life. Firstly, thanks to my supervisor Dr. Hannah Rohde for encouraging me throughout the entire process, even when it was out of both of our depths. My eternal gratitude goes to all of my participants who gave me their time despite no one understanding what my research was on. A special shout out to the 7th floor lounge of the Dugald Stewart Building which has been my second home these past few years. There's no place better for commiserating in good company.

And most of all thanks to all my friends who have been here for me and kept me going for the past four years. I couldn't imagine it without you and I honestly doubt I could have done it without all of you. Especially you Sherry, I'm so proud of both of us and I wouldn't be who I am now without you. Together we're slowly figuring it out. Now come along folks, this goes out to everyone who has been here and whatever comes next. Prost, 건배, Bersulang, Cheers!

Section B

Squibs & Write-ups

Critical Review of Bailyn (2003)

Qizhang Shi
University College London

Abstract. Bailyn (2003) discusses local-scrambling involving A-properties and long-distance scrambling involving A'-properties. Local scrambling is driven by the [D] feature on the head of IP and long-distance scrambling is triggered by a discourse-related [OP] feature. He then proposes a derivational system to describe the two types of scrambling without the reference to A and A'-properties. However, I criticise that the feature-driven analysis cannot capture the position of verbs regarding the adverb test in Russian OVS and the interpretive and formal license of Russian OVS order. By contrast, interface-based flexible-style analyses can, which assume some orders can be base-generated and make reference to the PF (phonological form) interface (Titov, 2021).

Keywords: syntax; Russian; local-scrambling; long-distance scrambling

---

1 Critical Review

Bailyn argues that long-distance scrambling involves A'-properties, which does not change binding relations.

(1) No binding effects
a.
Ja xoču, čtoby studenty$_i$ pročitali [knigi drug o druge$_i$]
I want that students read books-Acc about each other
'I want the students to read the books about each other.'

b.
[Knigi drug o druge$_i$]$_k$ ja xoču, čtoby studenty$_i$ pročitali t$_k$
books about each other I want that students read

Figure 1: *Bailyn, 2003, p. 43.*

A'-scrambling triggers weak crossover violations in (2b), when an object is scrambled over a coreferent pronoun. This is analogous to quantifier raising in (2a). This suggests again the A'-scrambled object reconstructs to its base position.

(2) Weak Crossover
a. *Ja xoču, čtoby ee sobaka poljubila každuju devočku
I want that [its$_i$ dog]-NOM loves [every girl]$_i$-ACC t$_i$
*'I want her dog to love every girl.
' b. *Každuju devočku ja xoču, čtoby ee sobaka poljubila
[every girl]$_i$-ACC I want that [its$_i$ dog]-NOM loves t$_i$
*'I want her dog to love every girl.'

Figure 2: *Bailyn, 2003, p. 44.*

Bailyn assumes Russian long-distance scrambling is discourse-driven triggered by the operator type feature of a functional category (Kitahara, 1997; Kawamura, 2001), making it semantically vacuous only with regard to binding effects but directly relevant for discourse interpretation. Conversely, local-scrambling exhibits A-properties. A-movement (subjecthood) tests show that all constructions in issue feed new binding relations and do not feed weak crossover violations.

The changes in binding relations can be tested using Principle A of the binding theory. In Figures 3-7, b sentences all permit the inverted constituent to bind an anaphor that it did not c–command before the movement.

(3)OVS

a. *[Otagai-no$_i$ sensei-ga] karera-o$_i$ hihansita
each other's teachers-Nom they-Acc criticized
*'Each other's$_i$ teachers criticized them$_i$.'

b. ?Karera-o$_i$ [otagai-no$_i$ sensei]-ga t hihansita
they-Acc each other's teacher-Nom criticized
'They$_i$ were criticized by each other's$_i$ teachers.'

Figure 3: *Saito, 2003, p. 485.*

(4)dative experiencers
a. ???Svoja$_i$ rabota nravitsja Maše$_i$
[self's work]-Nom pleases Masha-Dat
'Masha likes her work.'
b. Maše$_i$ nravitsja svoja$_i$ rabota
Masha-Dat pleases [self's work]-Nom
'Masha likes her work.'

Figure 4: *Bailyn, 2003, p. 50*

(5) possessive-PP inversion
a. ???[Svoj$_i$ dom] byl u Petrovyx$_i$
[self's house]-Nom was at the Petrovs
'The Petrovs had their own house.'

b. U Petrovyx$_i$ byl [svoj$_i$ dom]
at the Petrovs was [self's house]-Nom
'The Petrovs had their own house.'

Figure 5: *Bailyn, 2003, p. 50*

(6) iv. adversity impersonals
a. *Puljami prinadlezaščimi drug drugu ranilo milicionerov
bullets-INSTR belonging each other-DAT wounded the police-ACC
'Bullets belonging to each other wounded the policemen.

b. Milicionerov ranilo pujlami prinadlezaščimi drug drug
policemen-ACC wounded bullets-INSTR belonging each other-DAT
'The police were wounded by bullets belonging to each other.'

Figure 6: *Lavine & Freidin, 2002, p. 278.*

Similarly, a breach of Principle C occurs in Figures 7-11 when the inverted constituent binds an R expression that it did not previously c-command.

(7) i. OVS

a. [Novye znakomye Ivana$_i$] predstavili ego$_i$ predsedatelju
new friends of Ivan. introduced him-Acc chairman-Dat
'Ivan's$_i$ new friends introduced him$_i$ to the Chairman.'

b. *Ego$_i$ predstavili [novye znakomye Ivana$_i$] predsedatelju
him introduced new friends of Ivan chairman-Dat
*'He$_i$ was introduced to the chairman by Ivan's$_i$ new friends'

Figure 7: *Bailyn, 2003, p. 51.*

(8) a. [Sluxi ob Ivane$_i$] volnujut ego$_i$.
rumors-Nom about Ivan worry him-Acc
'Rumors about Ivan$_i$ worry him$_i$.'

b. *Ego$_i$ volnujut [sluxi ob Ivane$_i$]
him-Acc worry rumors-Nom about Ivan
*'He$_i$ is worried by the rumors about Ivan$_i$.'

c. [Ego$_i$ znakomyx] volnujut [sluxi ob Ivane$_i$]
[his friends]-Acc worry rumors about Ivan
'His$_i$ friends are worried by the rumors about Ivan$_i$.'

Figure 8: *Bailyn, 2003, p. 51.*

(9)
ii. Locative PP inversion

a. [Znakomye Ivana$_i$] byli u nego$_i$ doma.
friends of Ivan were at him at home
'Friends of Ivan's$_i$ were at his$_i$ house.

b. *U nego$_i$ doma byli [znakomye Ivana$_i$].
at him at home were friends-Nom of Ivan
'At his$_i$ house were friends of Ivan's$_i$.'

Figure 9: *Bailyn, 2003, p. 51.*

(10) iii. Possessive-PP inversion

a. ?[Igruški Ivana$_i$] byli u nego$_i$.
toys-Nom of Ivana were at him ?
'Toys of Ivan$_i$ he$_i$ had.'

b. *U nego$_i$ byli [igruški Ivana$_i$].
at him were toys-Nom of Ivan
*'He$_i$ had toys of Ivan's$_i$.'

Figure 10: *Bailyn, 2003, p. 52.*

(11) iv. Dative experiencers
a. [Znakomye Ivana$_i$] nravjatsja emu$_i$.
friends-Nom of Ivana like him-Dat
'Friends of Ivan$_i$ please him$_i$.' (cf *'He likes friends of Ivan.')

b. * Emu$_i$ nravjatsja [znakomye Ivana$_i$].
he-Dat like-pl friends-Nom of Ivan
*'He$_i$ is liked by friends of Ivan$_i$.'

Figure 11: *Bailyn, 2003, p. 52.*

Moreover, local-scrambling does not trigger weak crossover violations.

(12) a. *Ee$_i$ sobaka ljubit každuju devočku$_i$.
[her dog]-NOM loves [every girl]-ACC
'Her$_i$ dog loves every girl$_i$.'

b. [Každuju devočku]$_k$ ljubit ee sobaka t$_k$.
[every girl]$_i$-ACC loves [her$_i$ dog]-NOM
'Every girl is loved by her dog.'

Figure 12: *Bailyn, 2003, p. 52.*

(13) locative inversion

a. *[Ee uborščica] vošla [v každuju komnatu].
its cleaning lady entered into every room
'Its cleaning lady entered every room.'

b. V každuju komnatu vošla [ee uborščica].
into every room entered its cleaning lady
'Into every room entered its cleaning lady.'

Figure 13: *Bailyn, 2003, p. 52.*

(14) possessive-PP inversion
a. *[Ee$_i$ sobaka] byla na rukax u [každoj devočki]$_i$
its dog was on arms at every girl
'Her dog was in every girl's arms.'

b. ?U [každoj devočki]$_i$ byla na rukax [ee$_i$ sobaka]
at every girl was in arms her dog-NOM
'Every girl had her dog in her arms.'

Figure 14: *Bailyn, 2003, p. 53.*

(15) dative experiencers

a. ??[Ee sobaka] nužna [každoj devočke]$_i$
her dog-NOM needs every girl-DAT
'Her$_i$ dog is needed by every girl$_i$.'

b. [Každoj devočke]$_i$ nužna [ee sobaka]
every girl-DAT needs her dog-NOM
'Every girl$_i$ needs her$_i$ dog.'

Figure 15: *Bailyn, 2003, p. 53.*

Miyagawa (1997) suggests that local-scrambling is driven by the EPP. Languages that have V-to-I and morphological case marking allow the scrambling of objects due to the EPP. Bailyn extends EPP-analysis and unites all examples in Figures 3-4 and 16-22 as ‘Generalised inversion’.

(16) Russian -EPP acct of Adversity Impersonals

Soldata ranilo pulej
soldier-Acc wounded bullet-Instr
'The soldier was wounded by a bullet.'

Figure 16: *Lavine & Freidin, 2002, p. 256.*

(17) Locative Inversion (LI):

a. V klasse pojavilsja noven'kij PP-V-S
in class appeared new (one)
'A new boy appeared in class.'

b. Na posadočnuju polosu prizemlilsja samolet.
onto --------runway---------- landed airplane
'An airplane landed on the runway.'

Figure 17: *Bailyn, 2003, p. 48, citing Babyonyshev, 1996.*

(18) PP Inversion

a. U menja est' vopros. PP-V-S
at me is question-Nom
'I have a question.'

b. U nas rodilas' dočka PP-V-S
at us was born daughter-Nom
'A daughter was born to us

Figure 18: *Bailyn, 2003, p. 48.*

(19) Dative experiencers

a. Saše nravjatsja deti
Sasha-Dat likes-pl children-Nom
'Sasha likes children.'

b. Soldatam vidna doroga
soldiers-Dat visible-fem sg road-Nom-f-sg
'The soldiers can see the road.'

Figure 19: *Bailyn, 2003, p. 49.*

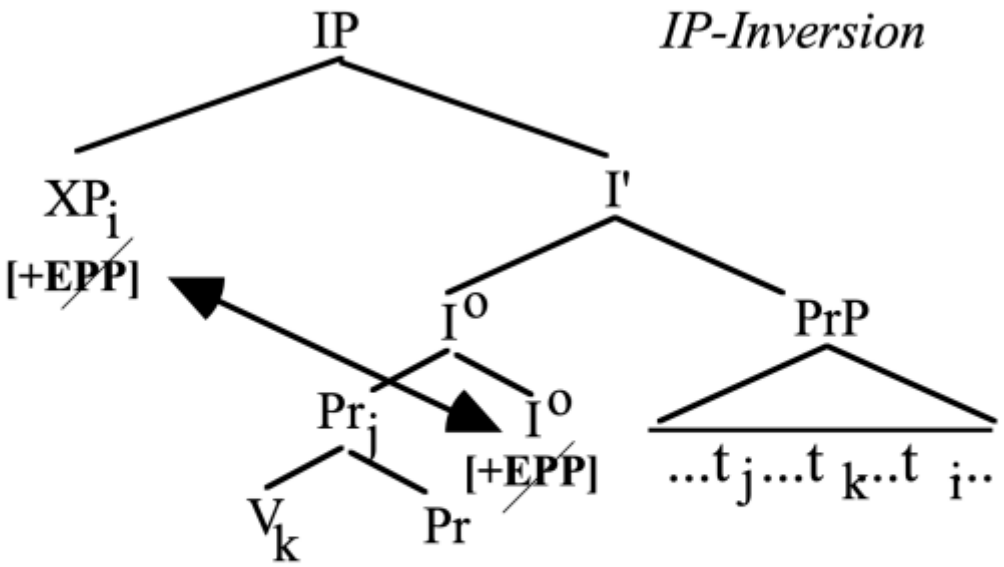
(20)
"Bad health" verbs
Menja tošnit ot ryby O-V-PP
me-Acc nauseates from fish
'I feel sick from the fish.'

Figure 20: *Preslar, 1998.*

(21) OVS
Étu knigu čitaet Ivan O-V-S
[this book]-Acc reads Ivan
'Ivan is reading this book.

Figure 21: *Bailyn, 2003, p. 49.*

(22) a. the schema of Generalized Inversion



b.characteristics of IP-inversion
-Non-Nominative XP in SpecIP:
-V comes before the subject;
-Differs from (Standard) Topicalization (IP-adjunction)

Figure 22: *Bailyn, 2003, p. 49.*

Examples that have the properties in Figure 22b are analysed as resulting from EPP-driven movement into SpecIP, with V-raising over the Nominative subject. This results in the order O/PP V-S/ Dat-V-Nom.

Bailyn proposes that the EPP is triggered by the strong [D] feature on the head of IP.

Additionally, Russian IP-inversion is followed by V raising to I triggered by a finiteness feature, which can be satisfied by either V-to-I or by the raised Nominative subject in SVO.

Bailyn proposes a derivational approach to represent the two types of scrambling while eliminating references to A or A'-movement and removing the need for reconstruction as a separate phenomenon. He adopts assumptions about derivational approach from Saito (2003):

i. Assume Copy Theory of Movement
ii. Assume XP arguments exhibit (at least) these features:
[P] (PF-relevant)
[D], [OP] (LF-relevant)
iii. Assume WH-movement and Long-Distance Scrambling are driven by [OP] feature
iv. Assume the motivation for Inversion (A-scrambling) is D-feature

Figure 23: *Some assumptions under the derivational approach.*

Bailyn adopts the 'feature-splitting' approach, allowing him to propose a purely derivational approach to scrambling. NPs are associated with LF and PF features, but features can be split by movement (Saito, 2003). Bailyn argues that NPs are interpreted and enter into binding relations at any point in the derivation where their [D] features are active. In Figure 24a EPP-driven scrambling, the XP moves with the [D] feature and the phonological feature. Since the [D] feature moves, the XP that undergoes this operation is interpreted and enters into binding relations in the landing site. Therefore, when the [D] feature raises, the EPP-driven movement can satisfy Principle A in Figures 3-6 and violate Principle C in Figures 7-11.

However, in Figure 24b discourse-driven scrambling, XP moves with the phonological feature. This movement is driven by the [OP], and the D feature remains in the base position. This is where the NP enters into the binding relation, and where it is reconstructed.

(24)
a. EPP-driven scrambling:(local, A)
[$_{IP}$XP$_i$ [D],[P] [$_{I'}$...t$_i$ [~~D~~],[~~P~~] ...]]

b. Discourse-driven Scrambling: (long, A')
[$_{IP}$XP$_i$ [P],[OP] [IP...t$_i$ [D],[~~P~~],[~~OP~~] ...]]

Figure 24: *Bailyn, 2003, p. 57.*

Although this feature-driven analysis seems to account for the examples above, I argue that it is problematic. In Bailyn's analysis, local-scrambling is driven by D-feature on the head of IP. This approach aligns with the Universal Base Hypothesis, which assumes that SVO is the universal base order and all alternative orders are derived by movement (Zwart, 1997). Bailyn argues that to derive Russian OVS, the direct object moves out of VP into the SpecIP, and the verb undergoes V-to-I, crossing the Nominative subject. In Russian, the [+T] feature on I can be satisfied by either movement of the subject to SpecIP (resulting in SVO), or by V-to-I movement, in which case a non-subject moves to SpecIP to satisfy the EPP (resulting in OVS). In English, there is no V-to-I movement available, so

English lacks the OVS order. Therefore, the [+T] must be satisfied by movement of the subject to SpecIP.

However, V-to-I analysis is not supported by adverb placement tests (Titov, 2013). In both SVO and OVS in Russian, the adverb marking the left margin of the VP must precede the verb, clearly indicating that the verb remains within the VP (Titov, 2013).

(25)
a. Ja dumaju, čto Ivan často celuet Mašu
I think that Ivan-NOM often kisses Masha-ACC
'I think that Ivan often kisses Masha.'
b. * Ja dumaju, čto Ivan celuet často Mašu
I think that Ivan-NOM kisses often Masha-ACC

Figure 25: *Titov, 2013, p. 38.*

(26)
a. Ja dumaju, čto Mašu často celuet Ivan
I think that Masha-ACC often kisses Ivan-NOM
'I think that Ivan often kisses Masha.'
b. * Ja dumaju, čto Mašu celuet často Ivan
I think that Masha-ACC kisses often Ivan-NOM

Figure 26: *Titov, 2013, p. 39.*

Moreover, Bailyn suggests that A'-scrambling is driven by discourse-related requirements; however, A-scrambling is purely triggered by a strong [+D] feature on the head of IP. He adopts the analysis of generalized inversion from Bailyn (2004) where he assumes Russian OVS is as economical as SVO because of the same number of operations. He also suggests that all movements involved in generalized inversion do not achieve interpretive effects, including OVS. However, OVS is interpretively licensed. OVS can only be used when the object is IS (Information-Structurally) prominent and the subject is IS non-prominent (Titov, 2021). While OVS order is ungrammatical in Figure 28 where both arguments are prominent and in Figure 29 where both arguments are non-prominent, OVS is possible in Figure 30 where the object is prominent and the subject is non-prominent.

(28) [What did the children do to the toy?]$_{CONTEXT}$
a. # Igrušku [SLOMALI]$_{FOC}$ deti OVS
toy.ACC broke children
b. Deti [SLOMALI]$_{FOC}$ igrušku SVO
children broke toy.ACC
'The children broke the toy.'

Figure 28: *Titov, 2021, p. 9.*

(29) [What happened?]$_{CONTEXT}$
a. # [Igrušku slomali DETI]$_{FOC}$ OVS
toy.ACC broke children

b. [Deti slomali IGRUŠKU]$_{FOC}$ SVO
children broke toy.ACC
'The children broke the toy.'

Figure 29: *Titov, 2021, p. 9.*

(30) [Who broke the/a toy?]$_{CONTEXT}$
Igrušku slomali [DETI]$_{FOC}$ OVS

toy.ACC broke children
'The/a toy was broken by (the) children.'

Figure 30: *Titov, 2021, p. 8.*

Furthermore, Bailyn's account fails to explain the formal license of Russian OVS. In this account, OVS can be derived by A-scrambling an object to SpecIP above the subject and should not rely on whether the object carries an accusative morphological case marker or not. It is not enough to say that OVS is possible in a language that has morphological cases, and therefore allows V-to-I. Even in such a language, OVS may not be possible if the object does not carry an accusative case marker in Figure 31 because the thematic relation is not formally represented. Only SVO interpretation is possible because this is represented by syntax: the agent is syntactically higher than the patient (Titov, 2013).

(31) [What's new with mother?]$_{CONTEXT}$
Mat' NAVESTILA DOČ' SVO/*OVS
mother-NOM/ACC visited daughter-NOM/ACC
'Mother visited daughter.'
'*Daughter visited mother.'

Figure 31: *Titov, 2013, p. 46.*

OVS is possible when the formal license is applied which means the thematic relations are morphologically recovered at PF (Titov, 2013).

(32) [What's new with mum?]$_{CONTEXT}$
Mamu NAVESTILA DOČ' OVS
mum-ACC visited daughter-NOM/ACC
'Daughter visited mum.'

Figure 32: *Titov, 2013, p. 46.*

There is an alternative interface-based analysis which allows to base-generate OVS (Titov, 2013). The

object is base-generated above the subject within VP and undergoes A-movement to SpecIP. Thus, the problem of ordering the subject and the verb in OVS disappears because it allows the verb to merge with the subject first. Under this analysis, OVS involves an inverse order of theta-role assignment, and is syntactically more costly (Titov, 2013).

Therefore, OVS requires an interpretative license. OVS is used if it maps transparently onto the discourse template in Figure 33 as in Figure 30 (Titov, 2012).

(33) Information Structure
ARGUMENT ARGUMENT
[+IS-prominent] >> [-IS-prominent]

Figure 33: *Titov, 2013, p. 36.*

Additionally, whenever scrambling representation is mapped on to PF, PF detects the inverse order of theta-role assignment and it needs to recover the thematic prominence relations in its own representation via inflectional morphology (e.g. m-case assignment algorithms) in Figure 32 (Titov, 2013).

2 Conclusion

In conclusion, Bailyn proposes a derivational approach to represent local-scrambling and long-distance scrambling. While local-scrambling is to satisfy the [D] feature on the head of IP, long-distance scrambling is driven by discourse-related [OP]. I criticize the feature-driven analysis on the basis of Russian examples: the position of verbs regarding the adverb in OVS and the interpretive and formal license of OVS. These issues are all accounted for in the interface-based analysis.

3 References

Babyonyshev, M. (1996). *Structural conventions in syntax and processing* (Doctoral dissertation, Massachusetts Institute of Technology).

Bailyn, J. F. (2003). A (purely) derivational account of Russian scrambling. In W. Brown, J. Y. Kim, B. H. Partee, & R. A. Rothstein (Eds), *Formal approaches to formal Slavic linguistics* (11th ed.) (pp. 41-62). Ann Arbor: University of Michigan Press.

Bailyn, J. F. (2004). Generalized inversion. *Natural Language & Linguistic Theory*, 22(1), 1-50.

Kawamura, T. (2001). *A feature-checking analysis of Japanese Scrambling* (Masters dissertation, Kanzan University).

Kitahara, H. (1997). *Elementary Operations and Optimal Derivations.* Cambridge, MA: MIT Press.

Lavine, J. E., & Freidin, R. (2002). The subject of defective T (ense) in Slavic. *Journal of Slavic linguistics*, 10(1-2), 251-287.

Miyagawa, S. (1997). Against optional scrambling. *Linguistic Inquiry*, 27(1), 1-26.

Preslar, M. (1988). The subject position in Russian impersonal sentences. *Linguistic Analysis*, 28(1-2), 33-66.

Saito, M. (2003). A derivational approach to the interpretation of scrambling chains. *Lingua*, 113(4-6), 481-518.

Titov, E. (2012). *Information structure of argument order alternations* (Doctoral Dissertation, University College London).

Titov, E. (2013). Scrambling and interfaces. In F. Bildhauer & M. Grubic (Eds), *Information*

Structure: Empirical Perspectives on Theory (pp. 35-55). Potsdam: Universitätsverlag Potsdam.

Titov, E. (2021). *Russian wh-questions*. Ms. UCL

Zwart, J. W. (1997). The Germanic SOV languages and the Universal Base Hypothesis. In L. Haegeman (Ed), *The New Comparative Syntax*, (pp. 246-267). London: Longman.