J Cult Cogn Sci (2024) 8:247–274
https://doi.org/10.1007/s41809-024-00149-3

RESEARCH PAPER

Phrasal frequency and literacy as predictors of individual differences in on-line processing and comprehension of english complex NP subject-verb agreement

Kinga Patterson · **James A. Street** · **Andriy Myachykov**

Received: 7 December 2023 / Revised: 21 August 2024 / Accepted: 14 September 2024 / Published online: 13 November 2024

Abstract We present experimental evidence suggesting that frequency and literacy predict online processing and comprehension of subject-verb agreement constructions by adult native speakers of English. We measured participants' eye fixations, reaction times, and response accuracy in a forced-choice task using audio-visual eye-tracking paradigm. Participants completed a battery of tasks, inc. the Literacy Rating Scale (Tarone et al., Literacy and Second Language Oracy-Oxford Applied Linguistics, Oxford University Press, 2013), Agreement Judgement Task (e.g., Veenstra et al., Frontiers in Psychology 5:783, 2014). The AJT involved matching an auditorily presented subject phrase to one of two images of easily distinguishable colours presented on a computer screen (e.g., stars, circles). Participants heard 42 test sentences, counterbalanced across the three types: Type 1 (e.g., '*The stars with the circles are blue*'), Type 2 (e.g., '*The star with the circles is blue*') and Type 3 (e.g., '*The star with the circles are blue*'*. Type 1 and Type 2 constructions are considerably more frequent in writing than in speech (Miller et al., Spontaneous spoken language: Syntax and discourse, Oxford University Press on Demand, 1998) with Type 2 producing more attraction errors (Bock et al., Cognitive Psychology 43:83–128, 2001; Becker, L., & Dąbrowska, E. (2020). Does experience with written language influence grammaticality intuitions? UK Cognitive Linguistics Conference: University of Birmingham [conference presentation].). Data were analysed with linear mixed effects models and generalised additive models. Results show lower literacy participants took longer to process sentential cues and made more attraction errors. These findings support usage-based research showing frequency and experience effects on online comprehension of canonical and non-canonical constructions (Farmer, T. A., Misyak, J. B., & Christiansen, M. H. (2012). Individual differences in sentence processing. In Cambridge handbook of psycholinguistics (pp. 353-364)., Street, Language Sciences 59:192–203, 2017), detection and production of agreement attraction errors (Becker, L., & Dąbrowska, E. (2020). Does experience with written language influence grammaticality intuitions? UK Cognitive Linguistics Conference: University of Birmingham [conference presentation].) and demonstrate how linguistic and attentional processes interact (Tomlin and Myachykov, Attention and salience, Handbook of Cognitive Linguistics, 2015). They also complement corpus-based studies by providing evidence that native speakers are sensitive to observed distributions (Miller et al., Spontaneous spoken

---

K. Patterson (✉) · J. A. Street
Department of Linguistics, Northumbria University, Newcastle Upon Tyne, UK
e-mail: kinga2.patterson@northumbria.ac.uk

A. Myachykov
Department of Psychology, Northumbria University, Newcastle Upon Tyne, UK

language: Syntax and discourse, Oxford University Press on Demand, 1998).

Keywords Usage-based · Subject-verb agreement · Eye-tracking · Literacy · Frequency

Introduction

Most of what is known about language processing is based on the data collected from highly educated, highly literate participants. However, research indicates that participants with lower educational attainment and lower levels of literacy are slower to make use of grammatical cues (Mishra et al. 2012) and less likely to make use of grammatical cues, particularly when processing non-canonical/less frequent constructions (Street 2020, 2017, Street & Dąbrowska 2014, Becker & Dąbrowska 2020). This indicates that frequency of encounter as well as type of linguistic experience lead to faster and more reliable processing of particular constructions and that literacy 'fine tunes' attention to grammatical cues when processing language.

Historically, researchers have been interested in native language users' production, processing and judgment of utterances in which the verb does not agree with the subject, thus violating posited subject-verb agreement 'rules' (see, e.g., examples 1-4, below). These so-called attraction 'errors' have been the focus of numerous empirical studies employing a variety of paradigms (Becker & Dąbrowska, 2020; Bock et al., 1991, 2001; Bock, Nicol, & Cutting, 1999; Haskell & MacDonald, 2005; Meyer & Damian, 2007; Staub et al., 2010; Veenstra et al., 2014). These studies have enabled researchers to study how conceptual information is mapped onto linguistic representation and have shown how the manipulation of grammatical and conceptual features of the subject phrase can affect processing of agreement.

However, only more recently have studies begun to investigate the relationship between group and individual differences in agreement skills with cognitive variables, such as Working Memory and Executive Function (e.g., Veenstra et al. 2014) and educational variables, such as educational attainment, level of literacy and print exposure (Mishra et al. 2012; Becker & Dąbrowska 2020). The aim of the present study was to further test the relationship between educational variables and the processing of complex subject NP constructions, to determine if agreement processing skills could be predicted by differences in language experience and literacy, offering a counterpart to production studies. To do this, we tested participants with varying levels of literacy and employed sentences with complex subject NP structures of differing frequencies in written contexts (Miller et al 1998). Following Veenstra et al.'s (2014) results indicating that variation in processing complex subject NPs is attributable to variation in cognitive variables (e.g., WM, EF), we tested both experiential variables (i.e., amount and type of linguistic experience) and cognitive variables (i.e., WM and EF) as key predictors of the processing of complex subject NP constructions.

Subject-verb agreement

Grammatical agreement refers to the relationship between words, phrases, and sentences which are compatible by virtue of morphological inflections carried by at least one constituent, such that the morpho/phonological features of one constituent, e.g., the verb, are related to morpho/phonological features of another constituent, e.g., the noun (Corbett, 2006, Hartsuiker et al., 2003). Grammatical agreement, particularly subject-verb agreement (SVA) is pervasive across most of the world's languages (Acuña-Fariña 2009) and generally encodes information related to person (1) number (2), gender (3) and case (4).

1. I am happy / She is happy.
2. The girl sings / The girls sing.
3. El chico es muy contento / La chica es muy contenta.
   [The boy is very happy / The girl is very happy.]
4. Der gute Mann / Des guten Mannes.
   [The good man / The good men.]

As can be seen, SVA can involve number, person, gender, and case together and can represent the same information on different constituents (e.g., determiners, nouns, verbs). In English, agreement is established in almost every sentence (Hartsuiker et al. 2003) and is one of the strongest grammatical cues that language users rely on for comprehension during continuous speech, making agreement one of the strongest cues with high validity in sentence processing (MacWhinney et al. 1984; Li & MacWhinney 2013). There are, however, competing theoretical and processing accounts of SVA in the language sciences.

Theoretical accounts of SVA

Formal syntactic accounts posit that phrases inherit agreement information from lexical heads. For example, in the Minimalist Program (Chomsky 2001), constituents are linked by a single syntactic operation, AGREE in which the features required by one element can be valued by matching features of a second element. First, AGREE identifies a constituent with an inherent feature (e.g., Nouns that are naturally specified for gender) and then matches this feature to another constituent (e.g., Verb) that does not inherently hold the same feature (Polinsky 2016). This operation ensures that generated sentences will only contain matched (valued) features reaching semantic interpretation. In turn, AGREE is also responsible for erasing uninterpretable morphosyntactic features (e.g., features that cannot be transferred or matched to the second element) because sentences with uninterpretable features cannot be passed on to the semantic contributions (Ristic 2020).

However, numerous counter examples resist a purely formal syntactic account. For example, 1-4 exhibit so-called attraction 'errors'. Attraction 'errors' occur when grammatical and notional numbers of the subject phrase do not align such that a grammatically plural subject can refer to a notionally singular subject (see 5.) and a grammatically singular subject can refer to a notionally plural meaning (see 6-8). These examples challenge formal syntactic accounts since the sentences should be 'uninterpretable' and yet they are deemed acceptable by many native speakers, leading to a debate as to whether such 'errors' can be considered ungrammatical, at least in English (see Becker and Dąbrowska 2020). Regardless, such attraction 'errors' suggest that semantic and other discourse factors as well as structural effects make agreement more or less likely (Brehm & Bock, 2013; Eberhard, 1999; Franck et al., 2006; Franck et al., 2008; Franck, Vigliocco, & Nicol, 2002; Gillespie & Pearlmutter, 2013; Solomon & Pearlmutter, 2004; Vigliocco et al., 1995).

5. King prawns cooked in chilli salt and pepper was very much better, a simple dish succulently executed. (Biber 1991 cited in Kim 2004)

6. The government are planning new tax increases. (Kim, 2004, p 1108)

7. The faculty are all agreed on this point. (Kim, 2004, p. 1108)

8. The committee haven't yet made up their mind/s (Bock et al. 1999)

By contrast, on several usage-based, cognitive, and construction grammar accounts, a language user's knowledge of grammar is captured by constructions (i.e., form-meaning pairings) rather than formal operations. These constructional schemas are abstracted from the events in which they manifest and can differ subtly in meaning, satisfying various semantic and discourse constraints. Meaning on these accounts is dynamic and equated with language users' conceptualisation of the world (Langacker, 1987a, 1987b, pp. 487–488) including whether NPs are construed as singular or plural entities. Grammatical distinctions are thus posited to be motivated by and sensitive to intended conceptualisations.

In Cognitive Grammar (e.g., Langacker 1987a, 1987b, 2008), agreement constructions are a specific case of conceptual overlap, representing the same information in multiple places (e.g., on determiners, nouns, verbs) in which agreement markers are meaningful in their own right. Although the meaning is schematic, agreement markers provide critical cues for language users to conceptualise meaning of utterances (MacWhinney 1984). In Radical Construction Grammar (e.g., Croft, 2001), agreement also has language specific semantic function in that it allows identification of roles that participants play in specific constructions with different languages marking different participants in semantically equivalent constructions. As can be seen by comparing the sentences in 9, below, in 'pro-drop' languages like Spanish, the participant role of the verb is solely expressed by the inflectional agreement marker and so formal agreement marking can be said to be semantically meaningful (Hoffmann, 2022, p. 56).

9. Compré un libro // Compró un libro

Buy-1PSing-PAST a-Masc. book-Masc. // Buy-3PSing-PAST a-Masc. book-Masc.

I bought a book // He bought a book.

The attraction 'errors' in 5-8 indicate that SVA is dependent on how language users conceptualise the Head NPs – either as whole entities or in terms of separate individuals, rather than purely formal, syntactic operations. In 5, 'king prawns' is construed as a whole dish, rather than individual prawns, and the verb 'to be' agrees with this construal. In 6-8. the government, the faculty and the committee, all singular collective NPs, are construed in terms of separate individuals and thus take plural verb agreement in accord with this construal. A similar analysis may also be applied to the case of the following pair:

10. This government has broken its promises.

11. This government have broken their promises. (Kim, 2004, p. 1113)

Here again the difference between the two sentences appears to lie in the way in which 'government' has been conceptualised by the speaker.

Usage-based accounts also posit an important role for frequency in language attainment and processing (see, e.g., Divjak 2019). On these accounts, language processing and attainment are strongly related to linguistic experience. Since the input that language learners are exposed to contains many recurrent patterns (i.e., specific forms are associated with specific meanings), learners can extract schemas capturing these patterns. Through repeated use, these form-meaning pairings become entrenched, and hence more easily accessible. Thus, according to usage-based accounts, more experience with a construction should result in a greater entrenchment and hence faster and more reliable retrieval during processing.

Processing accounts of SVA

Bock and Miller (1991) demonstrated that subject-verb agreement errors typically occur in sentences where the head and local noun mismatch in number (see example 12 below). In such cases the verb is 'attracted' to agree with a local noun (a noun that occurs between the head noun and the verb, here *cabinets*) rather than with the subject's head noun, *key*. Numerous empirical studies using sentence completion and sentence-picture matching tasks show that these types of SVA attraction 'errors' are sensitive to conceptual influences (e.g., Barker, Nicol, & Garrett, 2001; Bock and Eberhard 1993; Bock and Miller 1991; Bock, Nicol and Cutting 1999; Haskell and MacDonald 2005, Veenstra et al., 2014; Veenstra et al., 2015). For example, local nouns that are semantically related to the head noun, exert stronger attraction than those that are not related (Barker et al., 2001) or when the head noun is a collective noun which is grammatically singular, but with a plural notion, it is more vulnerable to plural attraction than when it is notionally singular (see 13-14).

12. The key to the cabinets are missing*.

13. The jury for the trials are missing*.

14. The judge for the trials are missing*. (all Barker et al., 2001)

However, whilst there is evidence that some speakers are more susceptible to attraction errors than others, most studies on subject-verb agreement are conducted on highly educated adults (and also on only a small subset of world's languages) with very few studies investigating the relationship between differences in agreement skills and other variables. Veenstra et al. (2014) investigated how Working Memory (WM) and Executive Function (EF) are related to the production of attraction errors using a picture description task, since any effects of WM in this methodology would be more directly related to agreement process than in sentence completion paradigm. Experimental materials consisted of one (or two) brightly coloured shape(s) on the left-hand side (e.g., circle or circles), and one (or two) smaller grey coloured shape(s) on the right-hand side (e.g., triangle or triangles). Participants were instructed to produce a sentence starting with the left object(s) (the head NP) followed by the right object(s) (the local NP), always using *next to,* in order to connect them and end with an inflected verb phrase that included the colour of the head noun.

Thus, the design of the task was such that the head noun could be singular or plural and could either match or mismatch the local noun, yielding grammatically correct sentences such as *'the circle next to the triangle is blue'* as well as mismatching constructions such as *'the circles with the triangle is blue*'*. The results revealed main effects of both WM (verbal and non-verbal) and EF showing that participants with higher WM scores made fewer agreement errors. This indicates that production of these complex subject NP constructions is a function of cognitive variables. However, since clauses with complex subjects with an intervening NP that takes over agreement are considerably more frequent in writing than in speech (cf. J. Miller and Weinert 1998: 135-143), it is possible that more experience with written language should also facilitate processing of these constructions (as acknowledged by Veenstra et al. 2014).

Being literate has a profound effect on speakers' mental grammars and there is a growing body of research showing education- and literacy-related effects in language acquisition and processing in children (e.g., Petersson et al., 2001; Reis and Castro-Caldas 1997) and adults (e.g., Dąbrowska, 2012, 2015; Dąbrowska, Pascual & Gomez-Estern, 2022; Dąbrowska & Street, 2006; Huettig & Pickering, 2019; Street & Dąbrowska, 2010, 2014, Street 2017, 2020). These individual differences were shown to be significant in processing speed (Farmer et al., 2012), lexical knowledge (Mulder and Hulstijn 2011), fluency (Clark et al., 2009) and morpho-syntactic knowledge (Brooks and Sekerina 2006; Dąbrowska and Street 2006, 2014; Street 2017). There is also evidence of education-related effects in the processing of complex subject NP constructions. Becker and Dąbrowska (2020) examined high and low academic attainment participants' acceptability judgements and production of grammatical and ungrammatical complex subject NP constructions (e.g., *the structure of the new buildings is fascinating / *the structure of the new buildings are fascinating*). The results showed that education has a substantial effect on the detection and production of agreement attraction for native speakers of English. The higher education attainment participants were significantly better at detecting agreement attraction errors than the lower education attainment participants and made fewer agreement attraction errors in sentence recall than low attainment participants and that agreement attraction errors are more easily detected in the written than in the spoken modality for high attainment participants. In a rating task of sentences with agreement attraction errors, high attainment participants accepted errors 54% of the time, as opposed to low attainment participants, who accepted agreement attraction errors 90% of the time. Becker and Dąbrowska (2020) showed that the written modality only facilitated the processing of such structures in the high attainment group.

Predictions

The present study will further test the relationship between the processing of complex subject NP constructions and educational and cognitive variables. Following Veenstra et al. (2014) we employed a VWP task to test participants processing of the following sentence types in an on-line agreement judgement task. Our predictions are as follows: First, with regard to frequency effects, we hypothesized that complex subject NPs that are less frequent in spoken language and more frequent in written contexts would take longer to process and result in higher occurrence of errors in comprehension. Second, with regard to effects of literacy, we hypothesized that since complex subject NPs are more frequent in written contexts, participants with higher levels of literacy would have a processing advantage compared to participants with lower levels of literacy because the latter will have had relatively less direct experience of complex subject NPs. Third, we also hypothesized that literacy level would predict processing efficiency. That is, participants with higher levels of literacy would spend less time fixating on the distractor in all sentence types (T1, T2, Ungrammatical). Fourth, following Becker & Dąbrowska's (2020) findings, we predicted that ungrammatical items would produce higher error rates. However, since the low academic attainment participants in that study showed high acceptance rates for ungrammatical complex subject NP constructions, we hypothesized that lower literacy participants in the present study would have similar processing time for both grammatical and ungrammatical items – i.e., there would not be an additional processing cost for lower literacy participants for ungrammatical items. Fifth, following Veenstra et al.'s (2014) results, we tested whether experiential variables (i.e., amount and type of linguistic experience) or cognitive variables (i.e., WM and EF) would be the key predictors of the processing of complex subject NP constructions. Following Chipere's (2001) findings, showing that improvement in the processing of complex NP constructions is a function of increased grammatical knowledge rather than merely increased WM capacity, as well as more recent work suggesting that WM is dependent on long-term representations for language mediated by language experience (e.g., Acheson et al. 2011, Jones & Maken, 2015), we hypothesised that whilst WM and EF would be key predictors of processing of complex subject NP constructions, language experience variables would remain a significant predictor of performance on the processing of complex subject NP constructions when cognitive measures (i.e., WM and EF) are added to our statistical model.

Method

Design

This study employed a within-subject design whereby participants interact with images in a screen-based workspace to perform a motor task. In this case, participants were prompted to choose between two images on screen by pressing the key corresponding to the correct image (left arrow key or right arrow key). Participants typically generate a saccade to the referent and keep fixating it until the final interpretation is selected. This response is then typically used for response-contingent analyses, for instance, overall looks to target over distractors or competitors (Salverda, Kleinschmidt and Tanenhaus, 2014).

Participants

Purposive sampling was used to recruit adult English native speakers. A priori power analysis using G*Power software (Faul et al., 2007) required 53 participants at 0.8 power. 52 participants responded positively to the recruitment and 48 participants met the inclusion criteria *(Mean age = 31.96, M = 23 &, F = 22, & DND = 3).* Demographic information such as employment status, education level, measured in number of years spent in compulsory and further education and socioeconomic status (SES), measured by the postcode of each participant according to the English indices of deprivation (Gov.uk, 2019) were collected. Participants received £10 compensation in the form of a high-street shopping voucher for their time.

Materials

Agreement judgement task

We created the Agreement Judgement Task (AJT), a computerised picture matching test based on Veenstra, Acheson and Meyer's (2014) paradigm, where participants had to match the sentence they heard to one of 2 images appearing on screen at the same time, revealing simple shapes (e.g., star, circle, square) with easily distinguishable surface features (colours e.g., blue, green, black). Only common nouns were used. Simple shapes were selected because they could easily be named in a picture description task, while surface features helped facilitate salience of the head noun within each trial (Veenstra et al., 2014). Each subject phrase consisted of a determiner and a head noun (singular or plural) followed by a preposition (with or next to), which is then followed by a determiner and a local noun (singular or plural), e.g., *The star with the circles is blue.* The complete list of sentences can be found in Appendix I. The pictures always consisted of one or two brightly coloured shapes, and one or two smaller, dissimilarly coloured shapes, with one of the images corresponding to the sentence played to participants, for example, *The stars with the circle are blue,* see Figure 1.

Sentence types were manipulated to include grammatical and ungrammatical items. The most frequent (in speech and writing) and least ambiguous items were Grammatical Type 1 items, e.g., *the stars with the circles are blue*, where the head noun, local noun and verb all match in number. Grammatical Type 2 items are the less frequent more ambiguous sentences, e.g., *the star with the circles is blue*, where the head noun and verb match in number but the local noun does not. Ungrammatical items, e.g., *the star with the circles are blue**, refer to sentences where the local noun and verb match in number but the head noun does not. The task was set up to measure participants' accuracy in matching the sentences to one of the visual arrays on screen on a trial-by-trial level. Response times were calculated from the onset of the visual stimuli until the decision (measured by key press) was made by the participant. Processing load was measured by summing the duration of all fixations on each of the visual matches, the incorrect (distractor) and correct (target) arrays, in each trial. The 42 experimental trials were randomised and counterbalanced, creating different versions of the same experiment.

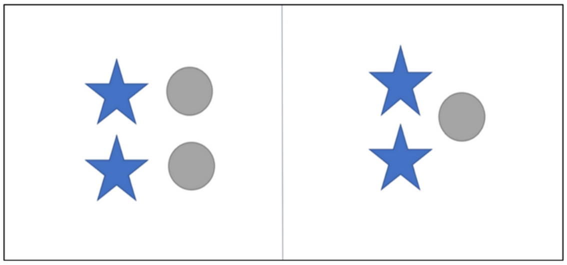


Fig. 1. Examples of visual displays of a single condition within the current Agreement Judgement Task

Literacy

The Literacy Rating Scale (LRS) employed here was developed specifically to measure adult native speakers' confidence in two basic literacy skills: reading and writing. The measure was adapted from Tarone et al. (2013) paradigm which they have used successfully in second language oracy studies with very low literate and illiterate populations. This scale was selected instead of the usually employed standardised literacy measures because standardised tests are often insensitive to individual differences and can potentially be aimed at a much higher level than low literate participants anticipate, making them uncomfortable or unwilling to continue. Participants were tasked with reading a short passage, offering a choice of 3 texts with different average sentence length and Flesh-Kincaid score (Counihan, 2021)[1] to suit the participants' literacy needs without compromising the integrity of the overall measure. Reading fluency, articulation, speed, and agility were rated.

Following this, a brief writing exercise was administered, measuring encoding skills such as grammar, punctuation, and vocabulary selection starting with 1–2-word answer items, gradually building up to questions requiring paragraph-length answers. Participants' attitude was also evaluated during the reading and writing tests, to gain some understanding of their confidence and surety of their own literacy skills. Tarone et al. (2013) study suggests that confidence in literacy skills can create a positive cycle, where individuals are more likely to seek out new reading materials, writing opportunities, and educational experiences, and the opposite effect would hold true for those with low confidence in their reading and writing skills. Those with confidence in their literacy skills would likely be approaching these tasks then as enjoyable experiences rather than daunting.

Education

Formal schooling has long been associated with the development of cognition. In modern societies, all children are required to attend compulsory education, and as a result, most children begin learning to read and write during their primary years. Years spent in education was measured using the UK education system, starting with Year 1 (children aged 5-6), going up in yearly increment, with the last option as a completed doctoral training, and ranged from 11-22 years (self-reported) in the sample. The mean number of years spent receiving formal education is approximately 15.5 years (SD = 2.68 years).

Cognitive variables

Working Memory (WM) has long been seen as a vital system within the human cognitive architecture that temporarily holds and manipulates information needed to perform cognitive tasks, such as sentence processing. WM was assessed here using a series of short Rapid Automatised Naming (RAN) tasks, a measure of Executive Function (EF), automatization, inhibition, and lexical access. RAN is the ability that combines the skills necessary for visual recognition and lexical processing with accurate and speedy speech production (Chiappe et al., 2002). Research indicates that RAN predicts several reading and phonological processing measures such as early-years' reading development (Schatschneider et al., 2002), vocabulary and letter knowledge (Peterson et al., 2017), short-term memory and speech production (Saletta et al., 2016), suggesting that learning to read promotes access to and retrieval of phonological representations (Araújo et al., 2019).

Arrays for the RAN used were adapted from Araújo et al. (2019) and were designed to reflect a variety of lexical items based on their imageability, familiarity, relative word-form frequency, word length and orthographic neighbourhood size (see Table 11 in Appendix III). Participants completed the same 4 arrays of different complexities, one after the other and were timed in each instance (measured in seconds). All 4 arrays were accomplished in under 4 minutes on average. To reduce the complexity of our data, a composite score was created in R (version 4.1.1; R Core Team 2022) using the 4 arrays of the RAN test,

---

[1] The current LRS offers the choice of 3 texts based on sentence numbers, average sentence length and Flesh-Kincaid score (Counihan, 2021). The lower the score, the more difficult it is to read with each level corresponding to an approximate age and school year. Text 1 in the current study has a Flesch-Kincaid score of 60.34F-K, meaning it should be easily understood by Key Stage 4 pupils around the age of 14-15. Likewise, Text 2 has a score of 73.95 F-K, and it should be easily understood by Key Stage 3 pupils around the age of 12-13. While Text 3 has a Flesch-Kincaid score of 85.13 F-K, it should be easily understood by Key Stage 2 pupils around the age of 10-11 and is representative of everyday conversational English.

weighted equally to represent an average performance across the different complexities.

Procedure

The study was approved by Northumbria University's Ethics Committee (submission ref 23033). Informed Consent was collected after allowing participants time to read through the Information Sheet and ask any questions that would not influence their performance on the tasks. Comfort breaks were offered at regular intervals to minimise fatigue.

Each trial consisted of the following procedure. Participants were seated in front of the screen, positioned approximately 65 cm (26") away from the display monitor (19" ViewSonic G90fb, 50–160 Hz frequency, display resolution 1024 x 768, sampling rate 1000 Hz). The EyeLink camera was adjusted to meet height requirements for each participant. Each recording session began with the 9-point calibration and drift correction. If the calibration error returned was more than 2.0, the process was repeated, making small adjustments on each turn. Instructions were presented in black, centred on a white background (font: Times New Roman, size 26).

After validation, participants completed 1 block of practice (6) and 2 blocks of experimental (42) trials of the Agreement Judgement task. Participants saw the Instructions on the display screen after calibration. To start the trials, participants had to press [SPACEBAR]. A blank screen then appeared for 300 msec and the onset of the display was timed at around 350 msec (target and distractor). Audio stimulus was played once at the same time as the visual stimuli appeared on screen. The objects remained in full view during the audio input and disappeared after the auditory stimulus was finished and the participant made a selection, measured by pressing either the [LEFT ARROW] or [RIGHT ARROW] keys on the keyboard. All other key presses were prohibited and discounted from the trials. After the selection, the trial ended and a blank screen was displayed for 500 msec, before the next trial began. The trial timed out after 10000 msec in case of 'no response', then moved onto the next trial. At the end of the session, participants were asked to complete literacy questionnaires, which were pen-and-paper based. Each testing session lasted approximately 45 minutes, including the practice blocks and comfort breaks, see Figure 2.

Sentences for the Agreement Judgement Task were recorded by a native Northern English speaker, using Audacity 3.2, a free, open source, cross-platform audio software (audacityteam.org 2022). Audio files were then edited, using the same software, to be presented at

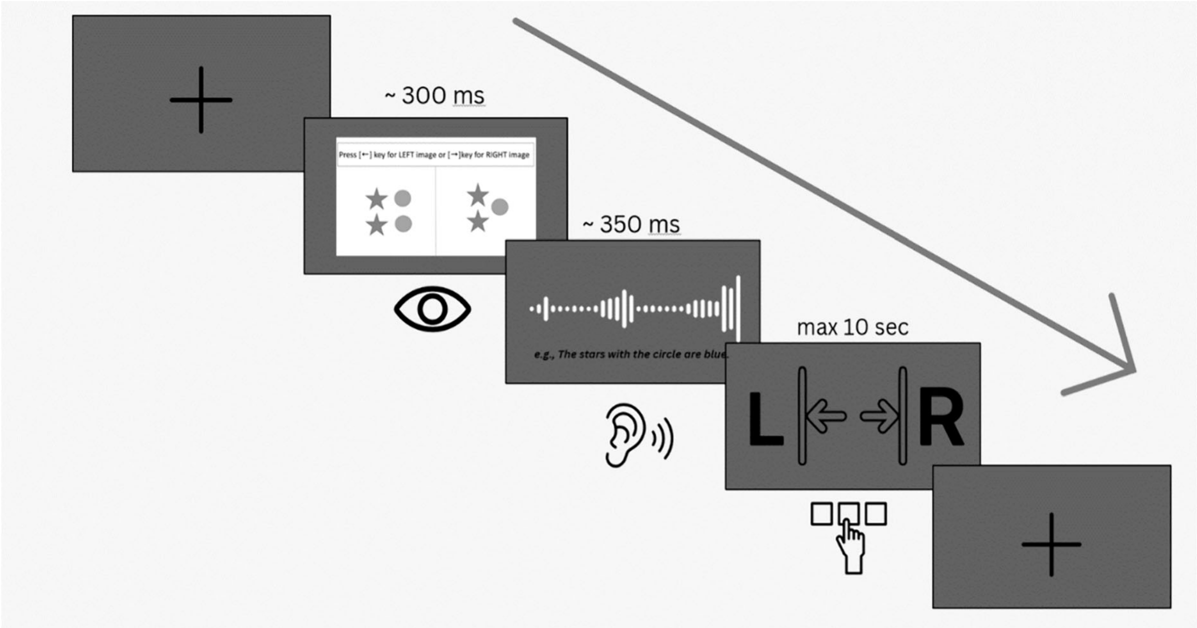


Fig. 2. A single trial within the current Agreement Judgement Task

the same frequency, sampling rate of 44000 and average length of 2.8 seconds.

The Agreement Judgement Task was configured in Experiment Builder (SR Research 2020). Participants' movements were recorded monocularly from the right eye using the EyeLink 1000 tower configuration with head support and chin rest at a sampling rate of 1000 Hz. The eye-tracking tasks were then deployed using SR research software (2020) and presented on a Windows 7 operating system. External headphones connected to the display computer were used to play the sentences throughout each recording session.

Data treatment

Pre-processing

Each recording session was processed using Data Viewer 4.2.1 (SR Research, 2020). Areas of interests (IA) were manually created around the distractor and the target images by drawing a rectangular shape the white background, closely matching the perimeter of each image (average size of IA: 234087 pixels). IAs were comparable in size (object sizes: 11 x 11 cm for single shapes, 13 x 8 cm for multiple shapes). Fixations outside of IAs and blinks were discounted from the analyses. Fixation and IA reports for each participant were generated with the following variables of interest: dwell time, number of fixations on target, number of fixations on distractors, and number of regressions between the target and distractor. Additionally, behavioural variables for key-press reaction times and accuracy were coded into the trial structure during the experiment building phase and downloaded post-recording (SR Research, 2020).

Data filtering and transformations

The data were aggregated at a participant-by-trial level. Analyses and transformations were completed in R studio version 4.1.1 "Kick Things" (R Core Team 2022). Practice trials were filtered out before data analysis. Reaction times were log transformed (Whelan 2008). We removed negative values (accidental key press before the stimulus onset), as it could be a sign of a misfired trial, values corresponding to reaction times that happened before the stimulus was heard (< 0.3 s), and time-outs (> 10 s). No further outliers were excluded based on SD, to avoid potentially distorting linear relations between RT and an independent variable (see Ulrich & Miller 1994 for a full discussion). This has resulted in a loss of 8% of the data, with 1864 observations in the final data set. Continuous variables, such as Literacy, were centred and scaled to help model interpretation.

Trial Type was coded to encompass SV phrase frequency and grammaticality. 'Grammatical Type 1' refers to the more frequent, less ambiguous phrases, e.g., *the stars with the circles are blue*, where the head noun, local noun and verb all match in number. 'Grammatical Type 2' are then the less frequent, more ambiguous phrases, e.g., *the star with the circles is blue*, where the head noun and verb match in number but the local noun does not. 'Ungrammatical Type' e.g., *the star with the circles are blue**, denotes sentences where the local noun and verb match in number but the head noun does not.

Observed power analyses

We conducted additional analyses to estimate the observed power in our models using bootstrapping methods, following the approach of Chernick and LaBudde (2014). Bootstrapping, while complex and not without limitations, provides an approximate measure of the observed power by repeatedly resampling the data and refitting the model to estimate the distribution of a statistic. We implemented the 'bootstrap power' function from the 'boot' package (Canty & Ripley, 2024; Davison & Hinkley, 1997) to perform this analysis, assessing the significance of predictors across resampled datasets.

Results

To test our key predictions, we investigated how phrasal frequency affected both mean reaction times and accuracy of identifying the correct visual match, and how literacy scores interacted with trial-type conditions. We also investigated how each trial type and literacy scores affected time spent looking at both the target and the distractor in each visual array. Distribution of the key variable was checked (see Figure 3 below).

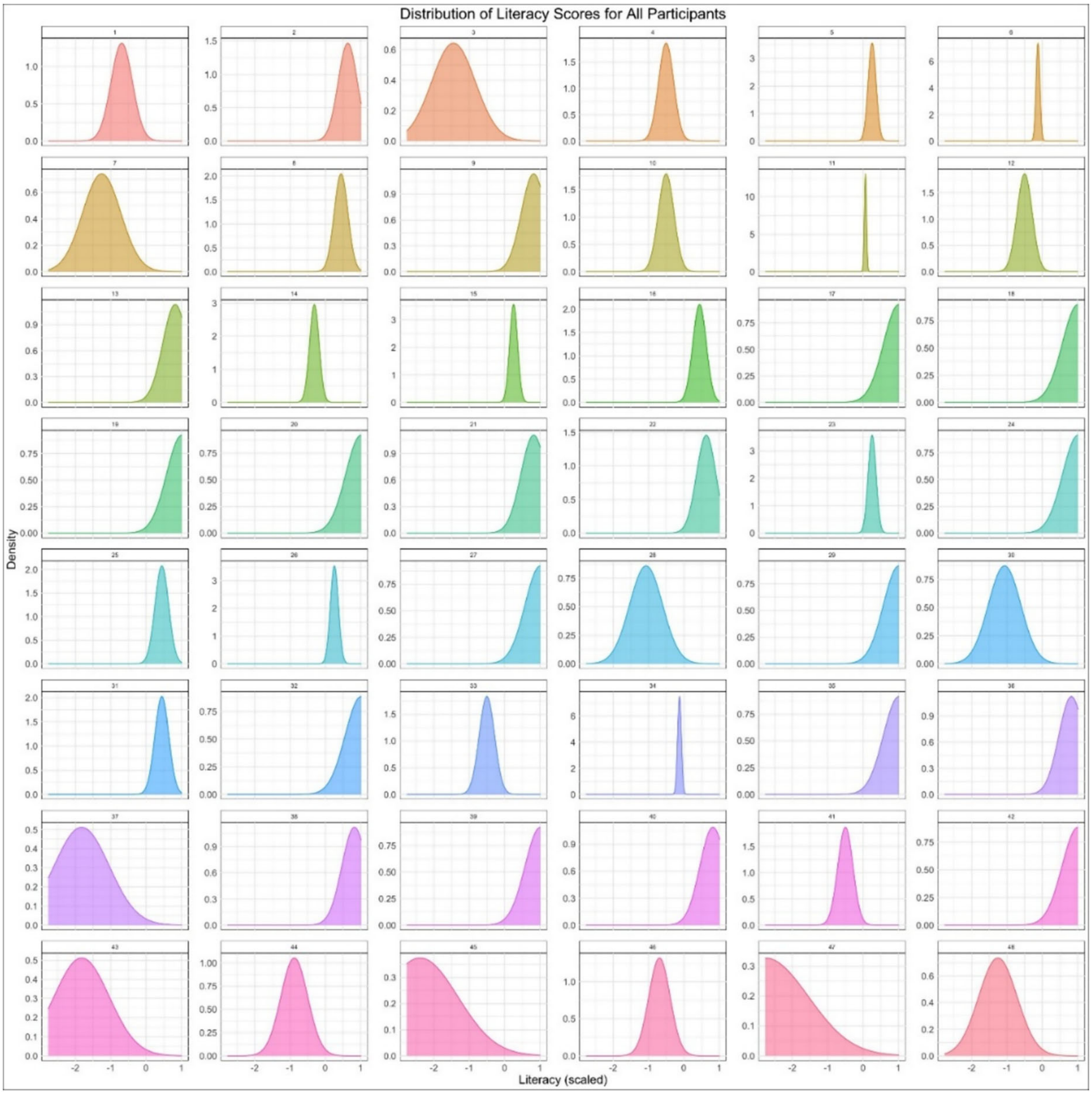


Fig. 3. Distribution of Literacy scores for all participants. Faceted density plots showing the distribution of scaled literacy scores (LRS_TOTAL_scaled) for individual participants. Each panel represents a separate participant, with the density of their literacy scores plotted in a distinct colour. The density plots are semi-transparent to allow for potential overlap in the distribution. The y-axis is scaled freely within each panel to accommodate differences in the distribution shapes

Response times

Log transformed response times were first grouped by trial types to get an overview of the whole sample. See Table 1.

The main effect of Trial Type is statistically significant (F(2, 1861) = 15.41, p < .001; Eta2 = 0.02, 95% CI [7.66e-03, 1.00]). Tukey's post-hoc tests revealed a statistically significant difference between grammatical type 1 and grammatical type 2 sentences (p < 0.001) and grammatical type 1 and ungrammatical sentences (p < 0.001), but not between grammatical type 2 and ungrammatical sentences (p = 0.94).

Table 1 Descriptive statistics for trial types

Trial type	N	M	SD	SE
grammatical type 1	640	8.17	0.25	0.01
grammatical type 2	619	8.24	0.25	0.01
ungrammatical type	605	8.24	0.26	0.01

N = Sample size, *Mean* = Mean score, *SD* = Standard deviation, *SE* = Standard error

This result is consistent with our predictions that type 2 sentences would take longer to process, see Figure 4.

Log transformed reaction times were also grouped by trial types and literacy groups to get an overview of how each group performed on average in each trial type. See Table 2.

Literacy groups performed similarly on average, with the High Lit Group being fastest in each trial-type condition. In ungrammatical and grammatical type 2 conditions, both Mid and High Lit groups performed similarly, see Figure 5.

To see if Response times can be predicted by our main targeted factors, we fitted an additive model (estimated using fREML and perf optimizer) first with smooth terms only (formula: Log_RT ∼ 1 + s(PARTICIPANT, bs = “re”) + s(Trial, bs = “re”)) to see whether our random effect candidates - participants and items - are justified (statistically). The dataset does not contain trial order so we cannot model random smooths of each participant across experimental trials. Nevertheless, both random effects are statistically justified, p = 0.003 and p < .001, respectively. Table 3.

The model’s intercept, corresponding to Literacy Scores = 0, Trial Type = grammatical type 1, Education = 0, is, at 8.08 (95% CI [7.99, 8.17], p < .001). Standardised parameters were obtained by fitting the model on a standardised version of the dataset. This model confirms our predictions and the initial overview that literacy and trial type are significant predictors of Response times in the Agreement Judgement Task. Education (t=2.001, p=0.046) is also statistically significant, however, the effect is smaller than that of Literacy (t= −3.030, p = 0.002). No interaction was observed for trial type and literacy scores for grammatical type 2, and ungrammatical sentences. Interactions between grammatical type 1 and literacy scores should be cautiously interpreted (within the Intercept) and thus, we will not be focusing on this further.

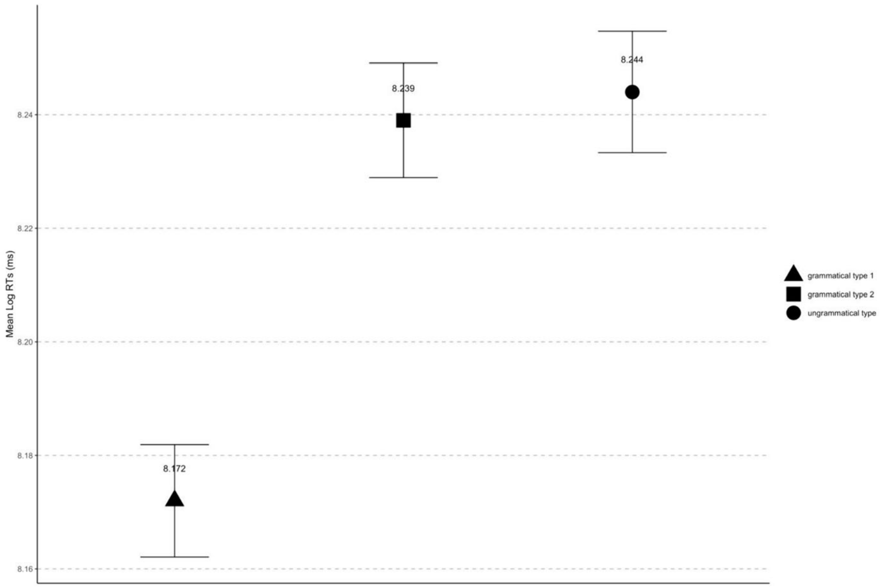


Fig. 4. Mean log response times in each trial type. Violin plot displaying the distribution of mean logarithmically transformed reaction times (Log RTs) across different trial types. Each violin plot represents the kernel density estimate of the data distribution for a specific trial type. The plot also includes error bars indicating the standard error of the mean, along with black-bordered points representing the mean Log RTs for each trial type. The shapes of the points vary by trial type, with corresponding shape and fill. The mean values are labelled above the points

Table 2 Descriptive statistics by literacy group and trial type

Literacy group	Trial type	M	SD	SE
Low Lit	type 1	8.27	0.23	0.03
	type 2	8.27	0.20	0.03
	ungrammatical	8.30	0.23	0.03
Mid Lit	type 1	8.22	0.27	0.03
High Lit	type 2	8.27	0.26	0.03
	ungrammatical	8.30	0.28	0.03
	type 1	8.15	0.25	0.01
	type 2	8.23	0.25	0.01
	ungrammatical	8.23	0.26	0.01

N = Sample size, *Mean* = Mean score, *SD* = Standard deviation, *SE* = Standard error

As described in 2.5.3., we used the bootstrapping method to estimate the observed power, due to the low sample size (Chernick and LaBudde, 2014). The power to detect the effect of Literacy is 78.2%, reflecting a fairly high likelihood of detection with our sample size. The power to detect the effect of Education is 53.9%, indicating a moderate likelihood of detection. As anticipated, the observed power for the interaction effects is low, at 10.2% and 8.2%, respectively; however, these interactions were not the primary focus of our current study.

Accuracy

We looked at the proportions of accurate responses for each trial types to show whether phrasal frequency affected the likelihood of choosing the correct visual match. We expected that more frequent items would be more accurately matched. See Table 4 below.

A Chi-square test indicated a statistically significant difference between proportions of accurate responses in each trial type (X = 215.44, df(2), $p < 0.001$), with all pairwise contrasts returning significant differences (all $p < 0.001$), see Figure 6.

We further analysed the proportions of accurate responses for each trial types in each literacy group to show whether literacy affected the likelihood of choosing the correct visual match in all sentence types. We expected that literacy would provide an advantage so the higher literacy participants would be more accurate in all conditions. See Table 5 below.

A Chi-square test indicated a statistically significant difference between proportions of accurate responses in each trial type (X = 236.158, df(2), $p < 0.001$), with pairwise contrasts between Low Lit and High Lit groups and Mid Lit and High Lit groups returning significant differences ($p < 0.001$, $p = 0.0017$, respectively). The contrast between Mid Lit and Low-Lit groups were found to be non-significant ($p = 0.131$).

Fig. 5. The relationship between literacy scores and trial types. Line plot showing the relationship between scaled literacy scores (LRS_TOTAL_scaled) and logarithmically transformed reaction times (Log RTs) across different trial types. The plot features three different trial types, each represented by a distinct line style: dotted, solid, and dashed. The lines indicate linear trends fitted to the data using a linear model (geom_smooth() with lm method). The colour of each line is determined by the trial type

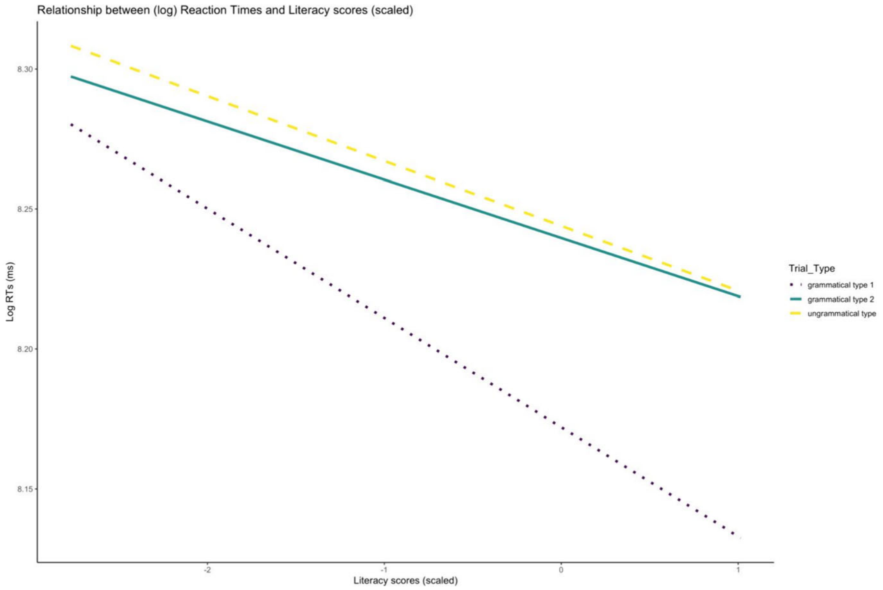


Table 3 Additive model predicting log transformed response times.

Effect	Estimate	*SE*	*t*	*p*
Fixed effects	−8.079	0.044	180.002	< 0.001 ***
Intercept	0.063	0.026	2.397	0.016 *
Type 2	−0.034	0.011	−3.030	0.002 **
Literacy RS	0.051	0.003	2.001	0.046 *
Education	0.070	0.026	2.649	0.008 **
Ungrammatical	0.014	0.002	0.920	0.358
Type 2 * Literacy	0.013	0.015	0.831	0.406
Ungrammatical*Lit				

Smooth terms	Edf Ref	F	*p*
Participant	0.889	9.760	< 0.002**
Item	43.552	1.846	< 0.001***

Log_RT ~ LRS_TOTAL_scaled * Trial_Type + EDU_YEARS + 1 + s(PARTICIPANT, bs = “re”) + s(Trial, bs = “re”). Continuous variables were centred and scaled; response times were log-transformed as detailed in Data Treatment. Significance codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘

Table 4 Proportions of accurate and inaccurate answers in each trial type

Trial type	Proportion of accurate responses	Proportion of inaccurate responses
grammatical type 1	86.25	13.75
grammatical type 2	85.78	14.22
ungrammatical type	74.05	25.95

Following Barr (2008), we fit a maximal random structure model using the Laplace Approximation but the optimization process for the model failed to reach convergence. We then simplified the model to remove items at the first instance, then when this still did not converge, we removed all random effects. The results of the final model can be found in Table 6 below.

As above, we used the bootstrapping method to estimate the observed power of the significant predictors in the model (Chernick and LaBudde, 2014). The power to detect the effect of Literacy is 86.8%, suggesting a high likelihood of detecting this effect with our sample size. The power to detect the effect of log-transformed response times is 97.7%, indicating a very high probability of detection. The power to detect the effect of ungrammatical sentence types is 100%, reflecting an extremely high likelihood of detection based on the observed data.

Interest area dwell time

Our dwell-time analyses (DT) focused on the time spent fixating the 2 main areas of interest defined in the experimental trials. Dwell time here is defined as the summation of the duration across all fixations on the given interest area. Table 7 summarizes the average DT values for each trial type.

Target DT did not significantly differ in each trial type (F(2, 1861) = 2.23, p = 0.108; Eta2 = 2.39e-03, 95% CI [0.00, 1.00]). However, the main effect of trial type in distractor dwell time is statistically significant (F(2, 1861) = 16.16, p < .001; Eta2 = 0.02, 95% CI[8.22e-03, 1.00]). Post-hoc revealed this difference is found between grammatical type 1 and grammatical type 2 sentences (p <0.0001), and ungrammatical and grammatical type 1 sentences (p < 0.001), but not between ungrammatical and grammatical type 2 sentences (p = 0.308), as seen in Figure 7.

These findings confirm that participants spent longer time fixating on the distractor in type 2 sentences and the least amount of time focusing on them in type 1 sentences. The extended time to decision could reflect a higher processing load in the less frequent, more ambiguous type 2 sentences. Interestingly, ungrammatical, and grammatical type 1 sentences were comparable when it came to fixating on the visual target. We also carried out similar analyses for each Literacy Group. The main effect of

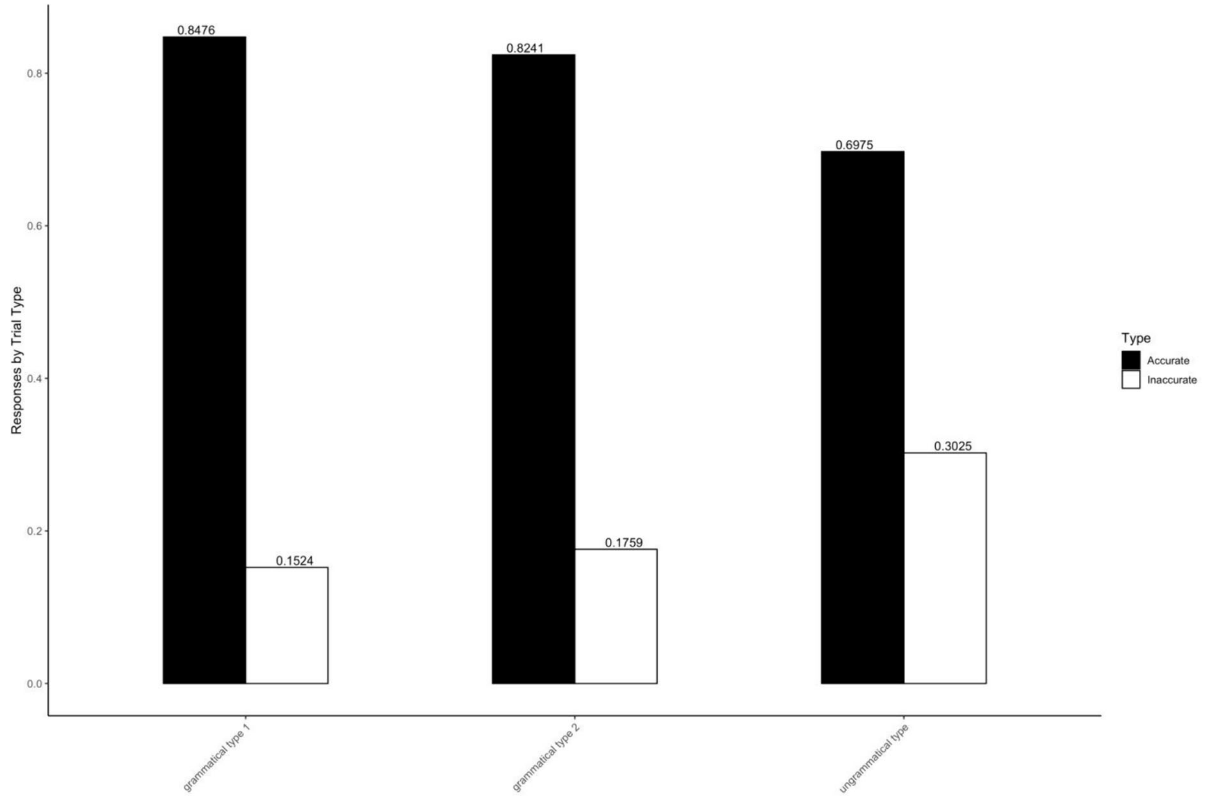


Fig. 6. Match accuracy for each trial type. Bar plots displaying the proportion of accurate and inaccurate responses across three trial types: "grammatical type 1," "grammatical type 2," and "ungrammatical type." The bars are grouped by response type (accurate vs. inaccurate) and presented side by side (dodge position) for each trial type. The black bars represent accurate responses, while the white bars represent inaccurate responses. Proportions are annotated above each bar. The x-axis labels are rotated for better readability

Table 5 Match accuracy for each trial type and literacy groups

Literacy Group	Proportion of Accurate Responses
Low Lit	67.53
Mid Lit	76.75
High Lit	84.96

Table 6 Logistic regression model predicting the likelihood of accuracy rates.

Effect	Estimate	*SE*	*z*	*p*
Fixed effects	-9.847	2.026	-4.860	< 0.001 ***
Intercept	0.020	0.169	0.115	
Type 2	-0.360	0.014	-3.154	0.908
Literacy RS	-0.005	0.028	-0.190	0.001**
Education	0.767	0.155	4.959	0.849
Ungrammatical	0.129	0.154	0.841	0.001***
Type 2 * Literacy	0.098	0.139	0.708	0.400
Ungrammatical*Lit	0.980	0.242	4.033	0.479
Log RT				< 0.001***

Result ∼ Trial_Type * LRS_TOTAL_scaled + EDU_YEARS + Log_RT, family = "binomial", data = data). Continuous variables were centred and scaled; response times were log-transformed as detailed in Data Treatment. Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '

target DT was significantly different in Literacy Groups (F(2, 1861) = 17.86, p < .001; Eta2 = 0.02, 95% CI[9.49e-03, 1.00]). Post-hoc revealed this difference is found between Mid and High Literacy Groups (p <0.0001), and Low and High Literacy Groups (p < 0.001), but not between Mid and Low Literacy Groups (p = 0.747). See Table 8.

The main effect of distractor dwell time was significant (F(2, 1861) = 38.55, p < .001; Eta2 = 0.04, 95% CI [0.03,1.00]). Post-hoc revealed this

Table 7 Interest area dwell time summary

Trial type	Target dwell time (ms)	Distractor dwell time (ms)
grammatical type1	1764.02	1054.06
grammatical type 2	1670.33	1320.06
ungrammatical type	1768.86	1248.06

difference is found between Mid and High Literacy Groups (p <0.0001), and Low and High Literacy Groups ($p < 0.001$), but not between Mid and Low Literacy Groups (p = 0.506)

Overall, as expected from the log transformed response times, the High Lit group spent the least amount of time fixating on all trial types. See Figure 8.

Cognitive and experiential variables

Table 9 shows that when the composite scores for Working Memory and Executive Function (indicated by Comp_WM in the table) are included in the model, both experiential variables and cognitive variables are significant predictors of accuracy, but that literacy shows a stronger association and a larger effect size.

AIC: 1677.9. The model's intercept, corresponding to Trial type = grammatical type 1, Literacy (scaled) = 0, Education (scaled) = 0, Log_RT = 0 and Comp_WM = 0, is at -10.49 (95% CI [−14.52, -6.50], $p < .001$).

Table 10 below further shows that both experiential variables and cognitive variables (as measured by the WM composite score) are significant predictors of RTs when included in the same predictive model. The WM tests were assessed in milliseconds, mirroring the measurement unit for RTs, which inherently contributes to their robust association.

Adj. R-sq. = 0.115, fREML = 73.561, n = 1864. The model's intercept, corresponding to Literacy (scaled) = 0, Trial_Type = grammatical type 1, Education (scaled) = 0, Comp_WM = 0, is at 7.92 (95% CI [7.80, 8.03], $p < .001$)

Discussion

The main purpose of the present study was to determine if individual differences in the processing

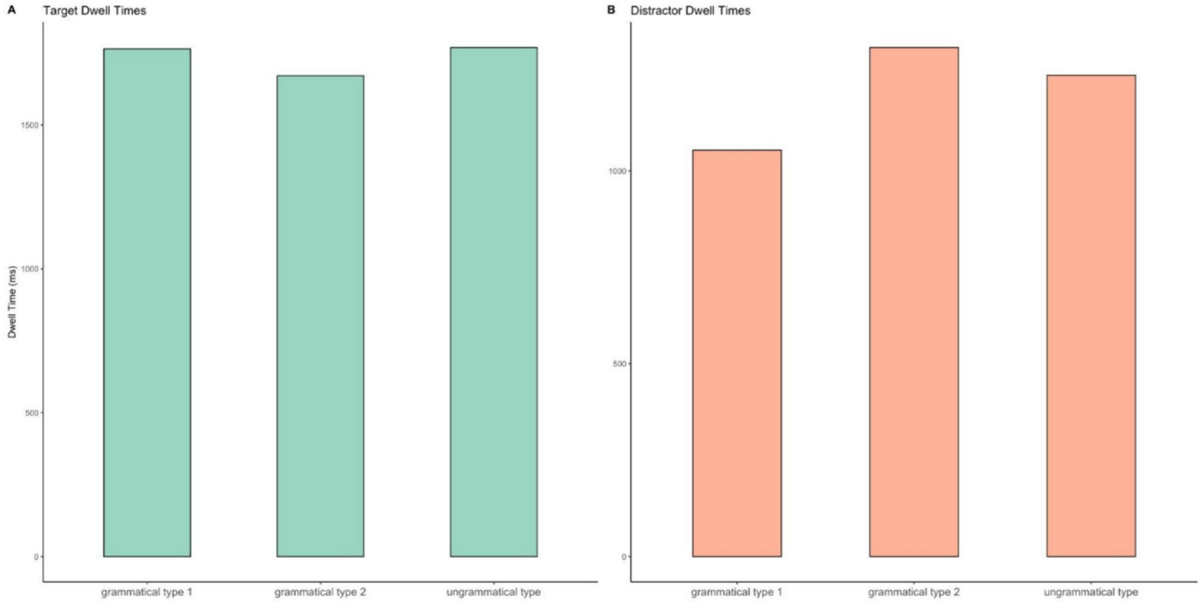


Fig. 7. Interest area dwell time in each trial type. Bar plots showing mean dwell times (in milliseconds) for target and distractor stimuli across three trial types: "grammatical type 1," "grammatical type 2," and "ungrammatical type." The left panel displays bar plots of target dwell times with a dashed line representing a linear trend, while the right panel shows bar plots of distractor dwell times with a similar trend line

Table 8 Interest area dwell time summary for each literacy group

Literacy Group	Target dwell time (ms)	Distractor dwell time (ms)
Low Lit	1996.42	1563.92
Mid Lit	1930.81	1470.86
High Lit	1661.42	1106.09

of complex subject NP constructions in attraction contexts could be attributed to differences in amount of linguistic experience and, following Mishra et al. (2012), literacy (as type of linguistic experience). We thus tested participants with varying levels of literacy and employed sentences with NP structures of differing frequencies in written contexts (Miller et al 1998). Following Veenstra et al.'s (2014) results indicating that variation in processing complex subject NPs is attributable to variation in cognitive variables, we tested whether language experiential variables (i.e., amount and type of linguistic experience) or cognitive variables (i.e., WM, EF) were key predictors of processing of complex subject NP constructions.

With regard to frequency effects, we hypothesized that complex subject NPs that are less frequent in spoken language and more frequent in written contexts would take longer to process and result in higher occurrence of errors in comprehension. The results confirmed this prediction. In our model, frequency of complex subject NP constructions significantly predicted response times and accuracy mismatch rates. Participants were significantly faster to process and made significantly fewer processing misinterpretations with the more frequent Type 1 sentences in which the 'intervening' local noun and the verb match in number (e.g., '*The stars with the circles are blue*'), than the less frequent Type 2 sentences in which the 'intervening' local noun and the verb do not match in number (e.g., '*The star with the circles is blue*'). With regard to effects of literacy, we hypothesized that

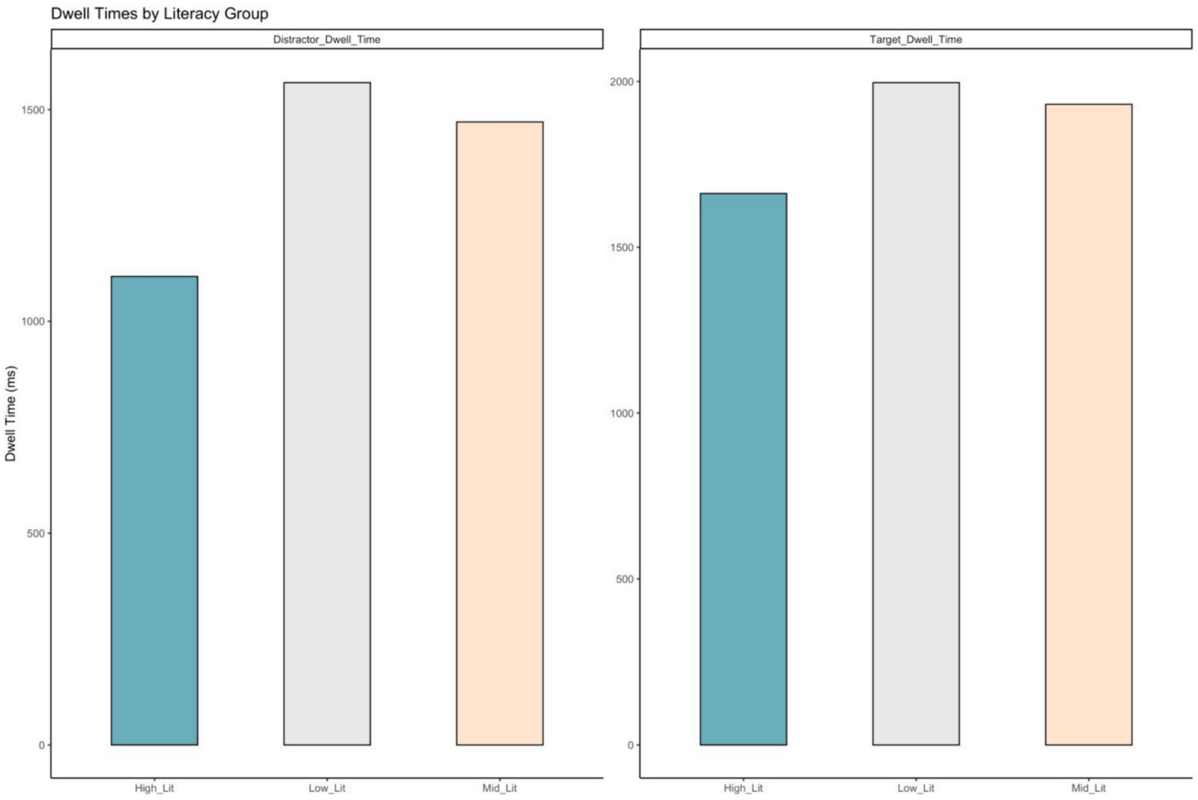


Fig. 8. Interest area dwell time in each literacy group. Bar plots showing mean dwell times (in milliseconds) for target and distractor stimuli across three literacy groups: Low Literacy (Low_Lit), Mid Literacy (Mid_Lit), and High Literacy (High_Lit). The left panel displays bar plots of target dwell times, while the right panel shows bar plots of distractor dwell times

Table 9 Predictive model for match accuracy with the addition of cognitive variables

Effect	Estimate	*SE*	*z*	*p*
Fixed effects	−10.490	2.026	−4.860	< 0.001***
Intercept	0.021	0.169	0.124	0.900
Type 2	−0.320	0.015	−2.773	0.005**
Literacy	−0.007	0.028	−0.274	0.784
Education	0.779	0.155	5.023	< 0.001***
Ungrammatical	0.129	0.153	0.844	0.400
Type 2 * Lit	0.101	0.139	0.728	0.466
Ungram*Lit	0.935	0.243	3.837	< 0.001***
Log RT	0.070	0.029	2.343	0.019*
Comp_WM				

since complex subject NPs are more frequent in written contexts, participants with higher levels of literacy would have a processing advantage compared to participants with lower levels of literacy because the latter will have had relatively less direct experience of complex subject NPs. This hypothesis was confirmed. Participants with higher levels of literacy were faster and produced fewer processing errors for all types of sentences (T1, T2, Ungrammatical) than the participants with lower levels of literacy.

We also hypothesized that literacy would predict processing efficiency. That is participants with higher levels of literacy would spend less time fixating on the distractor in all sentence types (T1, T2, Ungrammatical). This hypothesis was confirmed. Participants with lower levels of literacy spent on average 50% longer fixating on the incorrect visual match compared with participants with higher levels of literacy. With regard to grammaticality, we predicted that ungrammatical items would take longer to process. This hypothesis was confirmed. All participants took significantly longer to process ungrammatical items (the star with the circle(s) are blue*) compared to grammatical items (both T1 and T2 constructions). Following Becker & Dąbrowska's (2020) results, in which 90% of lower educated participants accepted attraction errors, we hypothesized that participants with lower levels of literacy in the present study would be likely to produce more processing errors with ungrammatical items but would not exhibit an additional processing cost, measured by dwell times. This hypothesis was not confirmed. Higher literacy participants were both faster at processing ungrammatical complex subject NP constructions and produced fewer processing errors that the lower literacy participants. We discuss this further, below.

Table 10 Predictive model for response times with the addition of cognitive variables

Effect	Estimate	*SE*	*t*	*p*
Fixed effects	7.9119	0.058	135.460	< 0.001***
Intercept	0.063	0.026	2.389	0.017*
Type 2	-0.028	0.011	-2.454	0.014*
Literacy	0.004	0.003	1.814	0.070
Education	0.070	0.026	2.667	0.008**
Ungrammatical	0.014	0.015	0.932	0.351
Type 2* Lit	0.013	0.015	0.862	0.389
Ungram*Lit	0.011	0.003	4.277	< 0.001**
Comp_WM				
Smooth terms	edf	Red.df	F	***p***
Participant	0.85	1	6.830	< 0.001**
Item	43.78	81	1.878	< 0.001***

Usage-based approaches

These results are consistent with usage-based accounts that predict a processing advantage for the most frequently occurring complex subject NP constructions for all participants since, on these accounts, repeated experience with particular constructions leads to greater entrenchment (Langacker 1987a, 1987b), which in turn results in faster and more accurate processing. The most frequently experienced structures then are accessed quicker and easier, becoming even more well-entrenched over time. On the other hand, constructions that are encountered less frequently, are likely to be embedded weakly (Brooks and Kempe 2019). Usage-based accounts also predict that this processing advantage should be more pronounced in higher literacy participants since these participants will have had relatively more direct experience of complex subject NP constructions. That is, the processing and interpretation of complex subject NPs should exhibit regularity x frequency x experience effects.

Regularity x frequency x experience

MacDonald and Christiansen (2002) posit that certain constructions are more regular (i.e., they have more consistent syntax to semantic mapping) and that sentence ambiguity resolution shows a frequency x regularity x experience interaction. On this view, interpretation of less regular, and less frequent sentence types (e.g., those with non-canonical syntax-meaning mappings) depends on direct specific experience (frequency of encounters) with that particular structure, which varies across individuals. Language users with more overall experience with language (e.g., highly skilled readers) consequently have more experience with both regular and irregular sentence types. However, the extra experience is most advantageous with 'irregular' forms. Variation in, for example, literacy/reading experience changes the nature of individual regularity x frequency x experience interaction for particular constructions.

Based on the underlying principle that 'singular subjects require singular verbs and plural subjects require plural verbs (Quirk et al, 1985)', in regular SVA constructions the grammatical and notional number of the subject phrase align without ambiguity (e.g., *the stars are blue, the star is blue, the stars with the circles are blue, the star with the circle is blue*). By contrast, less regular SVA constructions display non-canonical syntax- meaning mapping, where the grammatical and notional number of the subject phrase do not align, or there is ambiguity. For example, when subject is composed of two or more nouns or pronouns (*see 15*) or when the subject phrase includes a collective noun, which imply a plural entity but are considered singular and therefore take singular verb (*see 16*) or a more ambiguous syntax-semantic mapping e.g., when the subject NP and the verb are 'interrupted' by an intervening local noun (*see 17*).

15. The doctoral student and the committee members write every day.

16. King prawns cooked in chilli salt and pepper was very much better, a simple dish succulently executed. (Biber 1991 cited in Kim 2004)

17. The key to the cabinets is lost. (Bock et al 2001).

It is well-documented that in English, complex subject noun phrase structures are considerably less frequent than their less-complex counterparts (Miller, Miller and Weinert, 1998), with the mismatch effect occurring even less frequently in use. Thus, amount of reading experience has little effect on processing of regular constructions because speakers get lots of experience with these even if they read very little (or at all). However, literacy, and overall increased linguistic experience generally, affects processing non-canonical, more complex subject NP constructions because these are less frequent, have less reliable and/or more ambiguous syntax-semantic mapping and, thus, processing is dependent on direct experience with complex subject NP constructions. Therefore, one would expect faster processing and more reliable processing of frequent, regular/less ambiguous constructions, than less frequent, less regular, more ambiguous ones. Thus, one would expect complex subject NP constructions to be processed faster and more reliably by language users with more overall experience with language (e.g., highly skilled readers) because they are likely to have more direct experience with these constructions. The present study provides evidence of regularity x frequency x experience interaction for: complex subject NP constructions which support similar findings re: processing of relative clauses (MacDonald & Christiansen, 2002, Wells et al. 2009, Hutton and Kidd 2011) and other research indicating strong relationships between amount of reading and sentence processing and interpretation). Our results support MacDonald and Christiansen's (2002) proposition that frequency and experience effect the online comprehension of canonical and non-canonical constructions (Farmer et al. 2012, MacDonald & Christiansen 2002, Street & Dąbrowska 2014; Street 2017), as well as more recent research on native speakers' detection and production of agreement attraction errors (Dąbrowska & Becker 2020).

Type of linguistic experience as well as amount of linguistic experience is clearly important and being literate provides a different type of linguistic experience. One possible reason for this is that experience with decontextualised language helps to increase learners' metalinguistic awareness, which in turn leads language users to pay more attention to formal cues (see Street 2017 for further discussion). The current study has shown that not just the relative quantity of encounters matters (frequency of complex subject NP constructions), but also the quality. Those who read more encounter more written text and can make more accurate and faster decisions about linguistic choices.

Type of linguistic experience may help explain the results of the ungrammatical constructions. The higher literacy participants are faster and more accurate at processing because they are better at detecting grammatical cues. Participants with lower levels of literacy in the present study showed similar mismatch rates to the lower academic attainment (and plausibly lower level of literacy) participants in Becker and Dąbrowska's (2020) study. We, therefore, speculated that if lower academic attainment participants in Becker and Dąbrowska's study were more prepared to accept (and produce) 'ungrammatical' complex subject NP constructions as grammatical, the lower literacy participants in the current study might be faster processing 'ungrammatical' complex subject NP constructions. However, this was not confirmed.

It would appear that due to type of linguistic experience higher literacy participants encounter more tokens of complex subject NP constructions and also more verb types used in the construction. As a consequence, they develop more entrenched and abstract representation of the construction overall and this results in faster and more reliable processing. A related possibility is that participants have both grammatical and 'ungrammatical' complex subject NP schemas (see Fig 9 in Appendix II). However, the grammatical schema is more entrenched than the 'ungrammatical schema' for higher literacy participants in part because these participants encounter Type 1 and Type 2 sentence types more, but also because they are more likely to attend to grammatical cues and therefore have more experience of resolving cue type competition between, e.g., Head and Local NPs for verb agreement. Higher literacy participants still have a schema for 'ungrammatical' complex subject NP constructions which may account for why we do not find lower literacy participants processing Type 3 items faster as we speculated, but for lower literacy participants there is increased schema competition between grammatical and 'ungrammatical' schemas, and this has a processing cost.

With regard to processing accuracy, it is possible that because lower literacy participants are less likely to attend to grammatical cues, and because most every day spoken language has fewer complex constructions than written language, they are more likely to extract a simplex pattern of agreement from complex constructions i.e., they are more likely to 'extract' '*the circles are blue*' from '*the star with the circles are blue*' (see Fig 9 in Appendix II). There is evidence to support this claim. Dąbrowska (1997), for example, observed that the lower academic attainment adult participants in her study were much more likely than higher academic participants to 'extract' simple clauses (e.g., *Paul noticed the room was tidy*, *Shona was surprised*) from complex NP constructions (e.g., *Paul noticed that the fact that the room was tidy surprised Shona*).

These results are compatible with Langacker's account of 'blocking' (Langacker, 2017a, 2017b, 2008, p.235-237). Blocking occurs when a specific pattern takes precedence over a more general one, preventing the general pattern from applying in some particular situation because a more specific unit preempts it and reflects the built-in advantage of more specific units over more schematic ones in competing for activation. Since expressions with complex subject NPs such as '*The stars with the circles are blue*' and '*The star with the circles are blue*'.* are produced and accepted as conventional by native speakers of English, these language users will abstract the schema outlined in 18. in which the Head NP agrees with the verb and the Local NP can either agree with the verb (as in Type 1 sentences) or not (as in Type 2 sentences), as well as the schema outlined in 19. in which the Head NP does not agree with the verb but the Local NP does. Language users may also abstract a high-level schema outlined in 20. which represents some of the commonalities of the more specific lower level schemas.

18. $[\mathrm{NP}_{\mathrm{HEAD}}\ (+_{\mathrm{AGR}}\ \mathrm{V})\ \mathrm{NP}_{\mathrm{LOCAL}}\ (+_{\mathrm{AGR}}\ \mathrm{V}/{-_{\mathrm{AGR}}}\text{-}\mathrm{V})$ – $\mathrm{VERB}\ (+_{\mathrm{AGR}}\ \mathrm{NP}_{\mathrm{HEAD}})]$

19. $[\mathrm{NP}_{\mathrm{HEAD}}\ (-_{\mathrm{AGR}}\ \mathrm{V})\ \mathrm{NP}_{\mathrm{LOCAL}}\ (+_{\mathrm{AGR}}\ \mathrm{V})$ – $\mathrm{VERB}\ (+_{\mathrm{AGR}}\ \mathrm{NP}_{\mathrm{LOCAL}})]$

20. $[\mathrm{NP}_{\mathrm{HEAD}}\ (+_{\mathrm{AGR}}\ \mathrm{V})\ \mathrm{NP}_{\mathrm{LOCAL}}$ – $\mathrm{VERB}\ (+_{\mathrm{AGR}}\ \mathrm{NP}_{\mathrm{HEAD}})]$

Blocking can then account for the data on the basis that the lower level schemas, including the 'ungrammatical' schema, are available when processing. However, for the higher literacy participants the 'grammatical' subschema is more entrenched than the 'ungrammatical'. For the lower literacy participants, the 'ungrammatical' may be more entrenched than the 'grammatical' subschema or, as above, it may be the case that these participants are more likely to extract a simplex pattern, outlined in 21 (highlighted in grey shade in Figure 9) from a complex pattern – one that is captured by the more general and well entrenched schema for simplex subject NP verb agreement, outlined in 22.

21. [NP$_{\text{LOCAL}}$ (+$_{\text{AGR}}$ V) – VERB (+$_{\text{AGR}}$ NP$_{\text{LOCAL}}$)]

22. [NP(+$_{\text{AGR}}$ V) – VERB (+$_{\text{AGR}}$ NP)].

Based on ease of activation and degree of overlap with the target, the lower-level schemas 18. and 19. are activated when processing complex subject NP constructions. The high-level schema outlined in 20. (and indicated by a dashed-line box in Figure 9) is not available as it is blocked by the more specific units. Even if language users abstract the general pattern outlined in 20. in which the agreement of the Local NP with the verb is not specified, they do not necessarily exploit all the options it potentially makes available. It is the lower-level schemas not higher that actually determine what does and does not occur – including 'ungrammatical' constructions e.g., *The star with the circles are blue*.*

With regard to processing efficiency, our results are consistent with findings from other studies in which literacy predicted processing load in VWP and self-paced reading tasks (e.g., Mishra et al., 2012, Ashby et al., 2005). Mishra et al. (2012) suggested that this difference may be attributed to the notion that the acquisition and practice of reading and writing increases the likelihood of predictive processing i.e., from the grammatical cue to the 'target'. Individual differences in predictive behaviours have also been observed during reading (see Ashby, Rayner and Clifton 2005), it seems plausible that individuals with expertise in reading and writing are better able to generate expectations about future linguistic input (and outcomes) and launch saccades on the basis of these expectations. Our current findings support Mishra et al. (2012) and Ashby et al.'s (2005) results, demonstrating a close link between attention and measures of language and literacy skill (Tomlin & Myachykov 2015).

Language variables and cognitive variables as predictors of processing of complex subject NP constructions

Existing research (e.g., Veenstra et al. 2014; Veenstra et al. 2015) has demonstrated that cognitive variables affect the production of attraction errors. The findings showed main effects for both Working Memory (verbal and non-verbal) and Executive Function, demonstrating, for example, that children with higher WM scores made fewer agreement errors - presumably because they can keep more information in WM long enough for EF to determine that agreement with the Head NP is necessary and inhibit agreement with the Local NP (see, e.g., Engle et al., 2004), McCabe et al., 2010 for discussion on the interrelation between WM and EF). These results are consistent with other research showing that individual differences in cognitive variables, such as EF, predict many aspects of language processing in children and adults (Novick et al., 2014, Woodward et al., 2016, Nozari et al., 2016, Trude and Nozari 2017, Khanna and Boland, 2010, Festman et al., 2010).

However, there is also research which indicates that performance on cognitive variables, such as WM, is dependent on long-term representations for language mediated by language experience (e.g., Chipere, 2001, Acheson et al., 2011, Jones & Maken, 2015). With this in mind, we hypothesised that experiential variables, i.e., amount and type of language experience would remain a key predictor of processing of agreement in complex subject NP constructions even when WM measures are added to the model. The results of the present study support this hypothesis as well adding to findings from other research indicating that whilst cognitive variables predict many aspects of language processing, these cognitive abilities when applied for language processing are, in turn, mediated by and dependent on language experience.

Limitations

A limitation of the study is the small sample size. Testing hard-to-reach populations, such as those with low literacy or low educational attainment, poses a number of unique difficulties, including participant recruitment and retention and consequently small sample sizes are not unusual in such research. Mishra et al. (2012), for example, recruited 30 low literates in their study observing their eye-movements during gender agreement in sentential contexts. Dąbrowska et al. (2022) recruited only 15 semi-literate participants in their study investigating comprehension of Spanish object relative clauses. Similar to the present study, findings demonstrate clear effects of literacy in language processing and comprehension. Nevertheless, we have to acknowledge that such small sample sizes may not adequately represent the larger population from which they are drawn. As a result, the findings in the present study may not be generalisable beyond the specific individuals.

We recognise that our sample is relatively small compared to the mainstream individual differences studies; however, we possess significant data pertinent to individual differences, which is particularly valuable given the challenges associated with studying hard-to-reach populations. A notable aspect of our data is that the majority of individual differences studies on predictive language processing have been conducted with WEIRD populations, particularly undergraduate students, with some exceptions (e.g., Mishra et al., 2012, Dąbrowska et al., 2022 with low literacy adults; and Borovsky et al., 2012, and Nation et al., 2003 with children).

To further contextualise our findings, we compared our effect sizes with those from previous studies. While we already report partial eta squared values, we converted these to Cohen's D values for better comparability. Mishra et al. (2012) found small to moderate effect sizes in mean fixation proportions on the target object during both the initial (d = 0.23) and later time windows (d = 0.30). Our effect sizes for reaction times (RTs) between trial types are similar: between Grammatical Type 1 and Grammatical Type 2, d = 0.28, and between Grammatical Type 1 and Ungrammatical Type, d = 0.27. For dwell times, where we observed significant differences, the effect sizes were also comparable: Grammatical Type 1 vs. Grammatical Type 2, d = 0.31, and Grammatical Type 1 vs. Ungrammatical Type, d = 0.24. Unfortunately, Dąbrowska et al. (2022) did not report effect sizes, but they achieved significant results with only 15 participants. Therefore, our findings highlight how experiential variables influence the processing and comprehension of complex constructions in non-WEIRD populations.

Another possible limitation is the use of non-standardised literacy and cognitive measures with our current sample. Although standardised testing provides advantages such as the ability to easily compare scores to averages on a given metric, there are some widely attested biases in standardised tests (Reynolds & Suzuki, 2012) which can be insensitive to individual differences and can potentially be aimed at a much higher level than low literate participants anticipate, making them uncomfortable or unwilling to continue with testing. Furthermore, standardised measures are typically developed and validated on highly educated and Western populations, making them methodologically unsuitable to use with low literacy and socially disadvantaged participants (Henrich, Heine & Norenzayan, 2010).

Conclusion

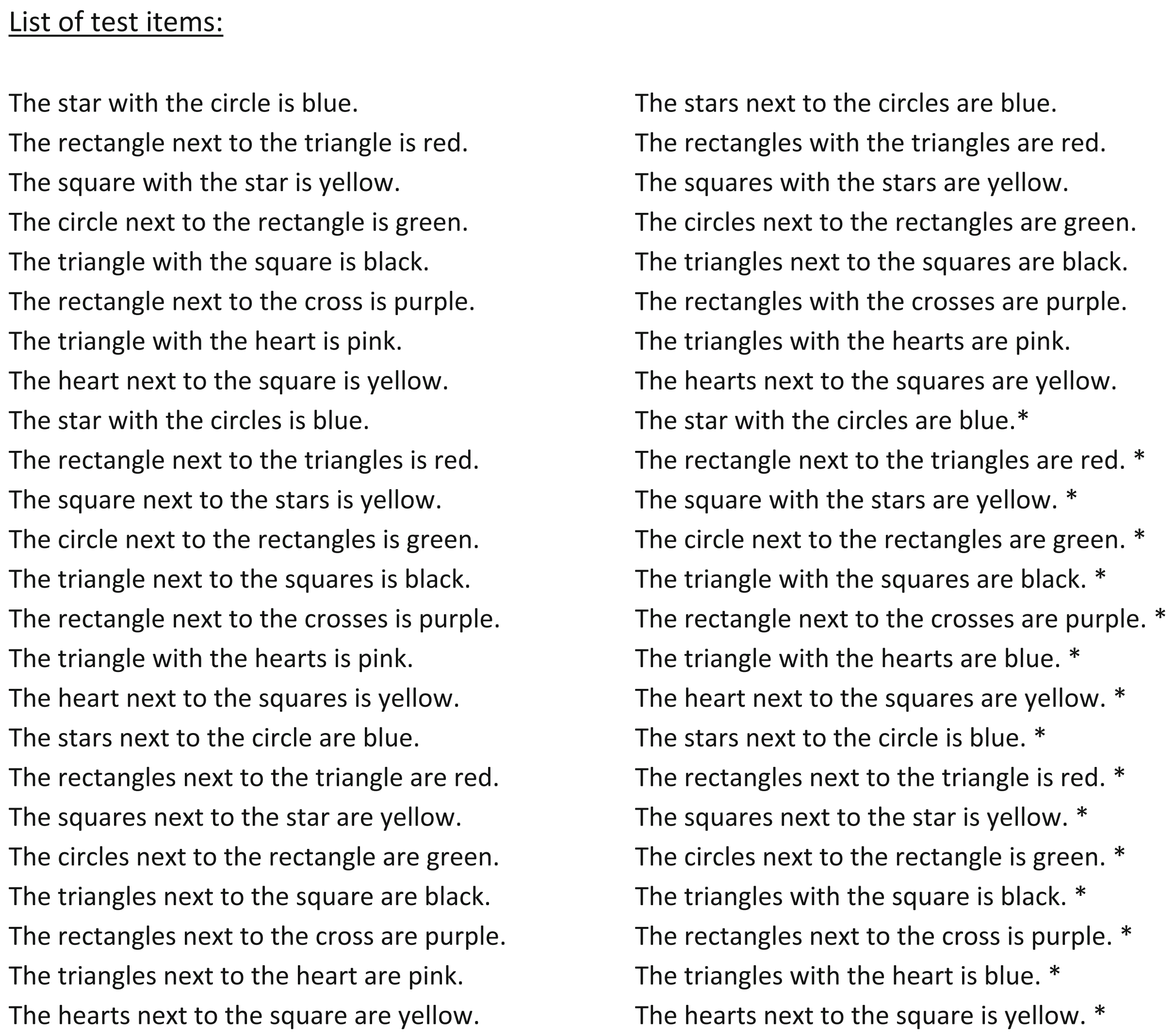
The results discussed in this paper provide new evidence that experiential variables effect the processing and comprehension of Complex Subject NP constructions. These findings converge with previous work investigating processing of agreement, and in particular agreement in complex subject NP constructions (e.g., Becker & Dąbrowska, 2020, Mishra et al. 2012, and Veenstra et al 2014) and provide support for usage-based approaches to language showing frequency and experience effects in the online processing and comprehension of canonical and non-canonical constructions (Farmer et al. 2012, MacDonald & Christiansen 2002, Street & Dąbrowska 2014; Street 2017). The study also adds to previous usage-based studies demonstrating how linguistic and attentional processes interact (Tomlin & Myachykov 2015), as well as complementing corpus-based studies by providing evidence from on-line processing that native speakers are sensitive to the observed distributions (Miller et al. 1998).

By contrast, the results are more problematic for formal, nativist accounts of grammatical competence since on these accounts innate grammatical knowledge, such as that which captures SVA (see Introduction, above) should not be predicted by experiential variables such as amount and type of language experience. Furthermore, in order to maintain the notion of an innate uniform syntactic competence, nativists have typically argued that any experiential-related individual differences in language processing and comprehension arise as a function of limited cognitive capacities (see, e.g., Chipere, 2001, 2003 for further discussion). However, the results of the present study challenge this account and provide support for approaches which posit that Working Memory abilities when applied to language are mediated by language experience.

In addition, the task design and methodology employed in the present study provide an effective alternative to traditional self-paced reading tasks for testing non-WEIRD populations. Testing low academic attainment, low literacy participants using literacy measures standardised on highly educated, highly literate populations is unrealistic. Furthermore, omitting the ‘decoding’ stage of reading, reduces the likelihood that the results of the present study are due to undiagnosed reading-related disorders (e.g., dyslexia, dysgraphia, orthographic processing disorder; Reilhac et al. 2012; Rothe et al. 2015).

The results are also consistent with results showing socioeconomic status (SES)-related differences in children’s processing of simplex and complex constructions – whilst processing of simplex constructions is similar for children regardless of SES background, there is a significant difference between SES groups in the processing of complex constructions with children from higher SES backgrounds outperforming their lower SES peers (Vasilyeva et al. 2008). However, whilst evidence indicates that lower SES children and adults have disproportionately poorer language (and educational) outcomes (Sirin 2005, Field 2010), SES is a distal predictor of these outcomes. More proximal predictors are amount and type of linguistic experience (Jago et al. 2018, Gilkerson et al. 2017, McGillion et al. 2017, Rowe 2012) and processing of complex language is a function of this: children from higher SES backgrounds typically and disproportionately have more exposure to, and more opportunity to produce more complex language. The results of the present study show that these amount and type effects on language processing and comprehension continue into adulthood and emphasise the significance of literacy (as a type of linguistic experience) on language development. They also highlight the importance of investigating individual differences in language processing and attainment as well as the need to diversify linguistic investigations beyond WEIRD populations (Henrich et al., 2010) in order to achieve a more comprehensive and inclusive understanding of language processing and comprehension.

Appendix I



List of test items:

The star with the circle is blue.
The rectangle next to the triangle is red.
The square with the star is yellow.
The circle next to the rectangle is green.
The triangle with the square is black.
The rectangle next to the cross is purple.
The triangle with the heart is pink.
The heart next to the square is yellow.
The star with the circles is blue.
The rectangle next to the triangles is red.
The square next to the stars is yellow.
The circle next to the rectangles is green.
The triangle next to the squares is black.
The rectangle next to the crosses is purple.
The triangle with the hearts is pink.
The heart next to the squares is yellow.
The stars next to the circle are blue.
The rectangles next to the triangle are red.
The squares next to the star are yellow.
The circles next to the rectangle are green.
The triangles next to the square are black.
The rectangles next to the cross are purple.
The triangles next to the heart are pink.
The hearts next to the square are yellow.

The stars next to the circles are blue.
The rectangles with the triangles are red.
The squares with the stars are yellow.
The circles next to the rectangles are green.
The triangles next to the squares are black.
The rectangles with the crosses are purple.
The triangles with the hearts are pink.
The hearts next to the squares are yellow.
The star with the circles are blue.*
The rectangle next to the triangles are red. *
The square with the stars are yellow. *
The circle next to the rectangles are green. *
The triangle with the squares are black. *
The rectangle next to the crosses are purple. *
The triangle with the hearts are blue. *
The heart next to the squares are yellow. *
The stars next to the circle is blue. *
The rectangles next to the triangle is red. *
The squares next to the star is yellow. *
The circles next to the rectangle is green. *
The triangles with the square is black. *
The rectangles next to the cross is purple. *
The triangles with the heart is blue. *
The hearts next to the square is yellow. *

Appendix II

See Fig. 9.

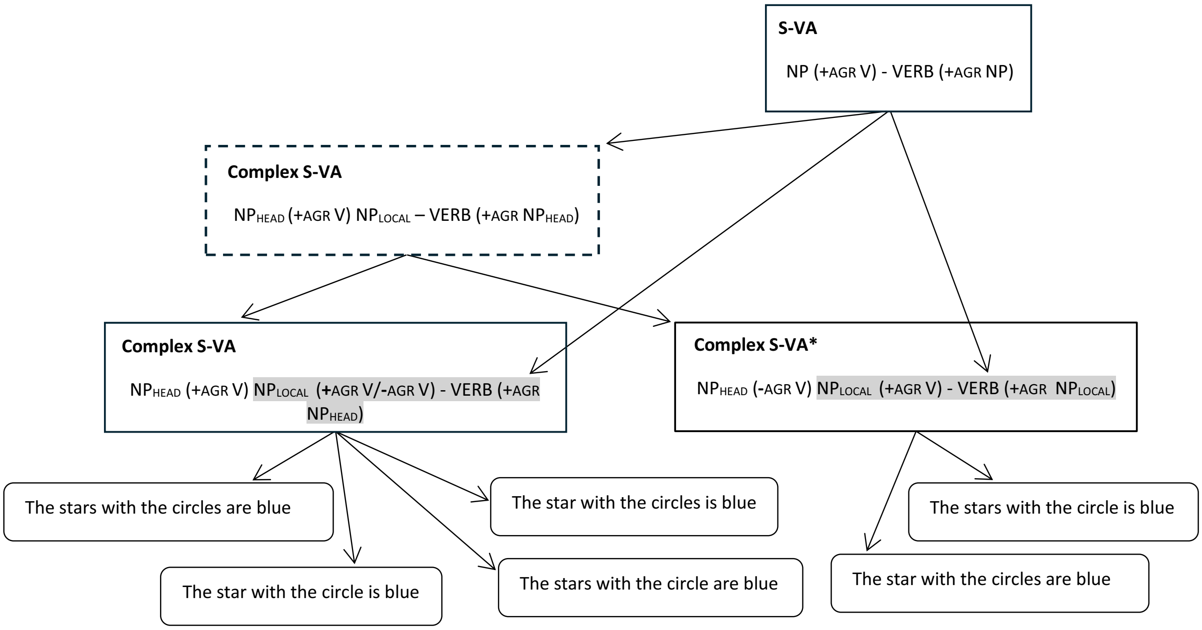


Fig. 9. Illustration of grammatical schemas in complex subject NP verb agreement

Appendix III

See Table 11

Table 11 Descriptive statistics for normative data for current version of RAN

	Shapes	Colours	Objects 1	Objects 2
Word length	5.2	4.6	6.2	6.8
M	0.8	1.1	2.6	1.3
SD				
Word Freq KF*	91.2	142.8	3.8	7.2
M	61.3	61.2	1.9	6.7
SD				
Ortho ND**	4.2	7.4	4.2	0
M	4	6	4.6	0
SD				

* Word frequency measures were calculated based on Kucera and Francis (1967) norms.

** Orthographic neighbourhood size density was calculated as the number of neighbours that look similar to the given cue word (Buchanan et al. 2019).

Author contributions All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Kinga Patterson. Technical expertise was provided by Andriy Myachykov. The first draft of the manuscript was written by Kinga Patterson and James A. Street, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding The work presented here was partially funded by the Northern Bridge Consortium, the corresponding author's doctoral training programme provider.

Data availability The datasets and R scripts that support the findings of this study are available from this anonymised OSF repository: https://osf.io/fmt9v/?view_only=17d1471f0ed64e05af2c96b7cc277a11

Declarations

Conflict of interest The submitted work is original and has not been published elsewhere in any form or language and has not been submitted for simultaneous consideration to any competing journals.

Ethical approval The study has been granted ethics committee approval prior to commencing from Northumbria University's Ethics Committee, submission reference 23033.

References

Acheson, D. J., MacDonald, M. C., & Postle, B. R. (2011). The effect of concurrent semantic categorization on delayed serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(1), 44.

Acuña-Fariña, J. C. (2009). The linguistics and psycholinguistics of agreement: A tutorial overview. *Lingua, 119*(3), 389–424.

Araújo, S., Fernandes, T., & Huettig, F. (2019). Learning to read facilitates the retrieval of phonological representations in rapid automatized naming: Evidence from unschooled illiterate, ex-illiterate, and schooled literate adults. *Developmental Science, 22*(4), e12783.

Ashby, J., Rayner, K., & Clifton, C., Jr. (2005). Eye movements of highly skilled and average readers: Differential effects of frequency and predictability. *The Quarterly Journal of Experimental Psychology Section A, 58*(6), 1065–1086.

Barker, J., Nicol, J. L., & Garrett, M. F. (2001). Semantic factors in the production of number agreement. *Journal of Psycholinguistic Research, 30*, 91–114.

Barr, D. J. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language, 59*(4), 457–474.

Becker, L., & Dąbrowska, E. (2020). *Does experience with written language influence grammaticality intuitions?* UK Cognitive Linguistics Conference: University of Birmingham [conference presentation].

Bock, K., & Miller, C. A. (1991). Broken agreement. *Cognitive Psychology, 23*(1), 45–93.

Bock, K., & Eberhard, K. M. (1993). Meaning, sound and syntax in English number agreement. *Language and Cognitive Processes, 8*(1), 57–99.

Bock, K., Nicol, J., & Cutting, J. C. (1999). The ties that bind: Creating number agreement in speech. *Journal of Memory and Language, 40*(3), 330–346.

Bock, K., Eberhard, K. M., Cutting, J. C., Meyer, A. S., & Schriefers, H. (2001). Some attractions of verb agreement. *Cognitive Psychology, 43*(2), 83–128.

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

Brehm, L., & Bock, K. (2013). What counts in grammatical number agreement? *Cognition, 128*(2), 149–169.

Brooks, P. J., & Sekerina, I. A. (2006). Shallow processing of universal quantification: a comparison of monolingual and bilingual adults. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, 28.

Brooks, P. J., & Kempe, V. (2019). More is more in language learning: Reconsidering the less-is-more hypothesis. *Language Learning, 69*(S1), 13–41.

Canty, A., & Ripley, B. D. (2024). boot: Bootstrap R (S-Plus) Functions. *R Package Version, 1*, 3–30.

Chernick, M. R., & LaBudde, R. A. (2014). *An introduction to bootstrap methods with applications to R*. John Wiley & Sons.

Chipere, N. (2001). Variations in native speaker competence: Implications for first-language teaching. *Language Awareness, 10*(2–3), 107–124.

Chiappe, P., Siegel, L. S., & Wade-Woolley, L. (2002). Linguistic diversity and the development of reading skills: A longitudinal study. *Scientific Studies of Reading*, *6*(4), 369–400.

Chipere, N. (2003). Understanding complex sentences: Native speaker variation in syntactic competence. Springer.

Chomsky, N. (2001). *Beyond explanatory adequacy*. MA: M.I.T.

Clark, R., Morrison, T. G., & Wilcox, B. (2009). Readers' theatre: A process of developing fourth-graders' reading fluency. *Reading Psychology, 30*(4), 359–385.

Corbett, G. G. (2006). Introduction: Canonical agreement. In *Agreement* (pp. 1–34). Cambridge University Press.

Counihan, B. (2021). An Open Educational Resource for Teaching Revision Flesch-Kincaid Readability Statistics. *HETS Online Journal, 11*, 1.

Croft, W. (2001). *Radical construction grammar: Syntactic theory in typological perspective*. Oxford University Press.

Dąbrowska, E. (1997). The LAD goes to school: A cautionary tale for nativists. *Linguistics, 35*(4), 735–766.

Dąbrowska, E., & Street, J. (2006). Individual differences in language attainment: Comprehension of passive sentences by native and non-native English speakers. *Language Sciences, 28*(6), 604–615.

Dąbrowska, E. (2012). Different speakers, different grammars: Individual differences in native language attainment. *Linguistic Approaches to Bilingualism, 2*(3), 219–253.

Dąbrowska, E. (2015). Individual differences in grammatical knowledge. In *Handbook of cognitive linguistics* (pp. 649–667).

Dąbrowska, E., Pascual, E., & Gómez-Estern, B. M. (2022). Literacy improves the comprehension of object relatives. *Cognition, 224*, 104958.

Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap Methods and Their Applications*. Cambridge University Press.

Divjak, D. (2019). *Frequency in language: Memory, attention, and learning*. Cambridge University Press. ISBN:9781107085756, 1107085756.

Eberhard, K. M. (1999). The accessibility of conceptual number to the processes of subject-verb agreement in English. *Journal of Memory and Language, 41*, 560–578.

Engle, R. W., & Conway, A. R. (2004). Working memory and comprehension. In *Working memory and thinking* (pp. 73–97). Routledge.

Farmer, T. A., Misyak, J. B., & Christiansen, M. H. (2012). Individual differences in sentence processing. In *Cambridge handbook of psycholinguistics* (pp. 353–364).

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191.

Festman, J., Rodriguez-Fornells, A., & Münte, T. F. (2010). Individual differences in control of language interference in late bilinguals are mainly related to general executive abilities. *Behavioral and Brain Functions, 6*, 1–12.

Field, F. (2010). *The Foundation Years: Preventing poor children from becoming poor adults The report of the Independent Review on Poverty and Life Chances*. The Stationery Office.

Franck, J., Vigliocco, G., & Nicol, J. (2002). Subject-verb agreement errors in French and English: The role of syntactic hierarchy. *Language and Cognitive Processes, 17*(4), 371–404.

Franck, J., Lassi, G., Frauenfelder, U. H., & Rizzi, L. (2006). Agreement and movement: A syntactic analysis of attraction. *Cognition, 101*(1), 173–216.

Franck, J., Vigliocco, G., Antón-Méndez, I., Collina, S., & Frauenfelder, U. H. (2008). The interplay of syntax and form in sentence production: A cross-linguistic study of form effects on agreement. *Language and Cognitive Processes, 23*(3), 329–374.

Gillespie, M., & Pearlmutter, N. J. (2013). Against structural constraints in subject–verb agreement production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(2), 515.

Gilkerson, J., Richards, J. A., Warren, S. F., Montgomery, J. K., Greenwood, C. R., Oller, D. K., Hansen, J. H. L., & Paul, T. D. (2017). Mapping the Early Language Environment Using All-Day Recordings and Automated Analysis. *American Journal of Speech-Language Pathology, 26*(2).

Hartsuiker, R. J., Schriefers, H. J., Bock, K., & Kikstra, G. (2003). Morphophonological influences on the construction of subject-verb agreement. *Memory & Cognition, 31*, 1316–1326.

Haskell, T. R., & MacDonald, M. C. (2005). Constituent structure and linear order in language production: evidence from subject-verb agreement. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(5), 891.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not WEIRD. *Nature, 466*(7302), 29–29.

Hoffmann, T. (2022). *Construction grammar: The structure of English*. Cambridge University Press.

Huettig, F., & Pickering, M. J. (2019). Literacy advantages beyond reading: Prediction of spoken language. *Trends in Cognitive Sciences, 23*(6), 464–475.

Hutton, J., & Kidd, E. (2011). Structural priming in comprehension of relative clause sentences. *The Acquisition of Relative Clauses: Processing, Typology and Function, 8,* 227.

Jago, L. S., Peter, M., Bidgood, A., Durrant, S., Pine, J. M., & Rowland, C. F. (2018). Individual differences in language acquisition: Identifying late talkers. In *Poster presented at the 3rd Lancaster Conference on Infant and Early Child Development (LCICD 2018).* Lancaster, UK.

Jones, G., & Macken, B. (2015). Questioning short-term memory and its measurement: Why digit span measures-long-term associative learning. *Cognition, 144,* 1–13.

Khanna, M. M., & Boland, J. E. (2010). Children's use of language context in lexical ambiguity resolution. *Quarterly Journal of Experimental Psychology, 63*(1), 160–193.

Kim, J. B. (2004). Hybrid agreement in English. *Linguistics*, *42*(6), 1105–1128.

Langacker, R. W. (1987a). *Foundations of Cognitive Gramma Theoretical Prerequisites* (Vol. 1). Stanford University Press.

Langacker, R. W. (1987b). *Foundations of Cognitive Grammar Descriptive Application* (Vol. 2). Stanford University Press.

Langacker, R. W. (2017). Cognitive Grammar. In B. Dancygier (Ed.), *The Cambridge Handbook of Cognitive Linguistics (Cambridge Handbooks in Language and Linguistics)* (pp. 262–283). Cambridge University Press.

Langacker, R. W. (2017). Conceptualization, symbolization, and grammar. *The new psychology of language* (pp. 1–39). Routledge.

Langacker, R. W. (2008). The relevance of Cognitive Grammar for language pedagogy. *Applications of Cognitive Linguistics, 9,* 7.

Li, P., & MacWhinney, B. (2013). Competition model. *The Encyclopaedia of Applied Linguistics, 1*, 1–5.

MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing Working Memory: Comment on Just & Carpenter (1992) and Waters & Caplan (1996). *Psychological Review, 109*(1), 35–54.

MacWhinney, B., Bates, E., & Kliegl, R. (1984). Cue validity and sentence interpretation in English, German, and Italian. *Journal of Verbal Learning and Verbal Behavior, 23*(2), 127–150.

McCabe, D. P., Roediger III, H. L., McDaniel, M. A., Balota, D. A., & Hambrick, D. Z. (2010). The relationship between working memory capacity and executive functioning: evidence for a common executive attention construct. *Neuropsychology, 24*(2), 222.

McGillion, M., Pine, J. M., Herbert, J. S., & Matthews, D. (2017). A randomised controlled trial to test the effect of promoting caregiver contingent talk on language development in infants from diverse socioeconomic status backgrounds. *Journal of Child Psychology and Psychiatry, 58*(10), 1122–1131.

Meyer, A. S., & Damian, M. F. (2007). Activation of distractor names in the picture-picture interference paradigm. *Memory & Cognition, 35*(3), 494–503.

Miller, J. E., Miller, J., & Weinert, R. (1998). *Spontaneous spoken language: Syntax and discourse*. Oxford University Press on Demand.

Mishra, R. K., Singh, N., Pandey, A., & Huettig, F. (2012). Spoken language-mediated anticipatory eye movements are modulated by reading ability: Evidence from Indian low and high literates. *Journal of Eye Movement Research, 5*(1), 1–10.

Mulder, K., & Hulstijn, J. H. (2011). Linguistic skills of adult native speakers, as a function of age and level of education. *Applied Linguistics, 32*(5), 475–494.

Nation, K., Marshall, C. M., & Altmann, G. T. (2003). Investigating individual differences in children's real-time sentence comprehension using language-mediated eye movements. *Journal of Experimental Child Psychology, 86*(4), 314–329.

Novick, J. M., Hussey, E., Teubner-Rhodes, S., Harbison, J. I., & Bunting, M. F. (2014). Clearing the garden-path: Improving sentence processing through cognitive control training. *Language, Cognition and Neuroscience, 29*(2),186–217.

Nozari, N., Freund, M., Breining, B., Rapp, B., & Gordon, B. (2016). Cognitive control during selection and repair in word production. *Language, Cognition and Neuroscience, 31*(7), 886–903.

Petersson, K. M., Reis, A., & Ingvar, M. (2001). Cognitive processing in literate and illiterate subjects: A review of some recent behavioral and functional neuroimaging data. *Scandinavian Journal of Psychology, 42*(3), 251–267.

Peterson, R. L., Boada, R., McGrath, L. M., Willcutt, E. G., Olson, R. K., & Pennington, B. F. (2017). Cognitive prediction of reading, math, and attention: Shared and unique influences. *Journal of Learning Disabilities*, *50*(4), 408–421.

Polinsky, M. (2016). Agreement in Archi from a minimalist perspective. *Archi Complexities of Agreement in Cross-Theoretical perspective, 1*, 184–232.

Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language*. Longman.

R Core Team. (2022). *R A language and environment for statistical computing*. R Foundation for Statistical Computing.

Reilhac, C., Jucla, M., Iannuzzi, S., Valdois, S., & Démonet, J. F. (2012). Effect of orthographic processes on letter identity and letter-position encoding in dyslexic children. *Frontiers in Psychology, 3*, 154.

Reis, A., & Castro-Caldas, A. (1997). Illiteracy: A cause for biased cognitive development. *Journal of the International Neuropsychological Society, 3*(5), 444–450.

Reynolds, C. R., & Suzuki, L. A. (2012). Bias in psychological assessment: An empirical review and recommendations. In *Handbook of Psychology* (2nd Ed., p 10).

Ristic, B. (2020). *Subject-verb agreement in real time: active feature maintenance as syntactic prediction* (Doctoral dissertation, Universidad del País Vasco-Euskal Herriko Unibertsitatea).

Rothe, J., Cornell, S., Ise, E., & Schulte-Körne, G. (2015). A comparison of orthographic processing in children with and without reading and spelling disorder in a regular orthography. *Reading and Writing, 28*, 1307–1332.

Rowe, M. L. (2012). A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. *Child Development, 83*(5), 1762–1774.

Saletta, M., Goffman, L., & Brentari, D. (2016). Reading skill and exposure to orthography influence speech production. *Applied Psycholinguistics, 37*(2), 411–434.

Salverda, A. P., Kleinschmidt, D., & Tanenhaus, M. K. (2014). Immediate effects of anticipatory coarticulation in spoken-word recognition. *Journal of Memory and Language, 71*(1), 145–163.

Schatschneider, C., Carlson, C. D., Francis, D. J., Foorman, B. R., & Fletcher, J. M. (2002). Relationship of rapid automatized naming and phonological awareness in early reading development: Implications for the double-deficit hypothesis. *Journal of Learning Disabilities*, *35*(3), 245–256.

Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research, 75*(3), 417–453.

Solomon, E. S., & Pearlmutter, N. J. (2004). Semantic integration and syntactic planning in language production. *Cognitive Psychology, 49*(1), 1–46.

SR Research Experiment Builder 2.3.38 [Computer software]. (2020). Mississauga, Ontario, Canada: SR Research Ltd.

Staub, A., White, S. J., Drieghe, D., Hollway, E. C., & Rayner, K. (2010). Distributional effects of word frequency on eye fixation durations. *Journal of Experimental Psychology: Human Perception and Performance, 36*(5), 1280.

Street, J. A., & Dąbrowska, E. (2010). More individual differences in language attainment: How much do adult native speakers of English know about passives and quantifiers?. *Lingua, 120*(8), 2080–2094.

Street, J. (2017). This is the native speaker that the non-native speaker outperformed: Individual, education-related differences in the processing and interpretation of Object Relative Clauses by native and non-native speakers of English. *Language Sciences, 59*, 192–203.

Street, J. A. (2020). More lexically-specific knowledge and individual differences in adult native speakers' processing of the English passive. *Language Sciences, 78*, 101254.

Street, J., & Dąbrowska, E. (2014). Lexically specific knowledge and individual differences in adult native speakers' processing of the English passive. *Applied Psycholinguistics, 35*(1), 97–118.

Tarone, E., Bigelow, M., & Hansen, K. (2013). *Literacy and Second Language Oracy-Oxford Applied Linguistics*. Oxford University Press.

Tomlin, R. S., & Myachykov, A. (2015). *Attention and salience*. Handbook of Cognitive Linguistics.

Trude, A., & Nozari, N. (2017, July). Inhibitory control supports referential context use in language production andcomprehension. In *Proceedings of the 39th Annual Conference of the Cognitive Science Society (CogSci)*.

Ulrich, R., & Miller, J. (1994). Effects of truncation on reaction time analysis. *Journal of Experimental Psychology: General, 123*(1), 34.

Vasilyeva, M., Waterfall, H., & Huttenlocher, J. (2008). Emergence of syntax: Commonalities and differences across children. *Developmental Science, 11*(1), 84–97.

Veenstra, A., Acheson, D. J., & Meyer, A. S. (2014). Keeping it simple: studying grammatical encoding with lexically reduced item sets. *Frontiers in Psychology, 5*, 783.

Veenstra, A., Meyer, A. S., & Acheson, D. J. (2015). Effects of parallel planning on agreement production. *Acta psychologica, 162*, 29–39.

Vigliocco, G., Butterworth, B., & Semenza, C. (1995). Constructing subject-verb agreement in speech: The role of semantic and morphological factors. *Journal of Memory and Language, 34*(2), 186–215.

Wells, J. B., Christiansen, M. H., Race, D. S., Acheson, D. J., & MacDonald, M. C. (2009). Experience and sentence processing: Statistical learning and relative clause comprehension. *Cognitive Psychology, 58*, 250–271.

Whelan, R. (2008). Effective analysis of reaction time data. *The psychological Record, 58*, 475–482.

Woodard, K., Pozzan, L., & Trueswell, J. C. (2016). Taking your own path: Individual differences in executive function and language processing skills in child learners. *Journal of Experimental Child Psychology,* 141, 187–209.