Bilingualism: Language and Cognition

cambridge.org/bil

Moses or Noah? A case of 'potato-potahto' when using a foreign language

Sara Dhaene[1], Nicolas Dirix[1], Hélène Van Marcke[1] and Evy Woumans[2]

[1]Department of Experimental Psychology, Ghent University, Henri Dunantlaan 2, 9000 Gent, Belgium and [2]Department of Translation, Interpreting and Communication, Ghent University, Groot-Brittanniëlaan 45, 9000 Gent, Belgium

Research Article

Cite this article: Dhaene S, Dirix N, Marcke HV, Woumans E (2022). Moses or Noah? A case of 'potato-potahto' when using a foreign language. *Bilingualism: Language and Cognition* **25**, 444–458. https://doi.org/10.1017/S1366728921000882

Received: 24 September 2019
Revised: 2 September 2021
Accepted: 7 September 2021
First published online: 9 November 2021

Keywords:
semantic illusions; anomaly detection; language processing; bilingualism; eye tracking

Address for correspondence:
Evy Woumans, Department of Translation, Interpreting and Communication, Ghent University, Groot-Brittanniëlaan 45, 9000 Gent, Belgium
evy.woumans@ugent.be

Abstract

Research among bilinguals suggests a foreign language effect for various tasks requiring a more systematic processing style. For instance, bilinguals seem less prone to heuristic reasoning when solving problem statements in their foreign (FL) as opposed to their native (NL) language. The present study aimed to determine whether such an effect might also be observed in the detection of semantic anomalies. Participants were presented NL and FL questions with and without anomalies while their eye movements were recorded. Overall, they failed to detect the anomaly in more than half of the trials. Furthermore, more illusions occurred for questions presented in the FL, indicating an FL disadvantage. Additionally, eye movement analyses suggested that reading patterns for anomalies are predominantly similar across languages. Our results therefore substantiate theories suggesting that FL use induces cognitive load, causing increased susceptibility to illusions due to partial semantic processing.

Introduction

Linguistic communication can be seen as a crucial characteristic of human functioning, being nearly indispensable in almost all aspects of daily life. In order for language to be effective, it is important that messages are grammatical and factual. However, this is not always the case. At times, interlocutors may inadvertently produce errors rendering their utterances nonsensical, such as when your neighbour says it *literally* took him a million years to get home, or when a friend talks about how they prefer Data to Spock in *Star Wars* (when these characters actually appear in *Star Trek*). Illogical as these statements may be, their erroneous nature is frequently unnoticed by the receiver, despite the fact that they possess the requisite knowledge to detect the error. Hence, this phenomenon has been referred to as 'knowledge neglect' (Marsh & Umanath, 2013). It was the aim of the current study to investigate whether processing such anomalies in a foreign language (FL) differs from processing them in the native one (NL).

In laboratory settings, knowledge neglect is often investigated by means of semantic illusions, in which a certain term has been replaced by a semantically related (but incorrect) term, causing incongruence within the sentence context. Illusions occur despite the fact that the anomalous word is being properly encoded by the receiver (i.e., heard or read; Erickson & Mattson, 1981; Reder & Kusbit, 1991), the receiver possesses the essential knowledge needed to detect the anomaly, and the receiver is warned in advance that some information will be ill-formed (Erickson & Mattson, 1981). In the literature, the inability to detect an anomalous utterance is often referred to as the *Moses illusion*. This name originates from a study in which participants were asked the following question: "How many animals of each kind did Moses take on the Ark?" (Erickson & Mattson, 1981). With some knowledge of the biblical story, it should be clear that the question is meaningless, since not Moses but Noah built an ark against the flood. Still, the majority of participants in the study answered with 'two'; the answer that would have been correct if the query comprised the *proper* proper noun 'Noah'.

In an attempt to explain the locus of the Moses illusion, Reder and Kusbit (1991) have formulated their Partial Match theory. This theory assumes that subjects correctly encode distorted sentences, but fail to completely match memory representations. As a query or statement is read, the terms or concepts are matched to memory so that an appropriate response may be retrieved. However, not every word or concept will be matched exactly to a corresponding memory trace. Instead, a more effortless process of partial matching will be carried out, which will allow distortions to slip by unnoticed as long as they bear sufficient resemblance to the intended word, as is the case with semantically (Van Oostendorp & De Mul, 1990) and phonologically (Shafto & MacKay, 2000) related words. This theory is corroborated by eye tracking research, which shows that only detected anomalies result in processing difficulties when reading distorted sentences (e.g., Bohan & Sanford, 2008; Cook, Walsh, Bills, Kircher & O'Brien, 2018). In other words, no slowdowns are observed for reading undetected anomalies, suggesting that they are also not implicitly detected. In contrast,

CAMBRIDGE UNIVERSITY PRESS

when anomalies are consciously detected, eye tracking patterns reveal increased fixations to the distorted word and longer total reading time (Bohan & Sanford, 2008; Daneman, Lennertz & Hannon, 2007). Still, first fixation reading times for detected and undetected anomalies appear to be equal. This corresponds well to the findings of studies into pragmatic anomalies and inconsistencies that are readily detectable (e.g., Braze, Shankweiler, Ni & Palumbo, 2002) and into more demanding reading comprehension (e.g., Weiss, Kretzschmar, Schlesewsky, Bornkessel-Schlesewsky & Staub, 2017).

Although the Moses illusion seems quite robust (e.g., Barton & Sanford, 1993; Hannon & Daneman, 2001; Umanath, Dolan & Marsh, 2014), there are a few moderating factors that alter the magnitude of the effect. For instance, expertise does not completely nullify the illusionary effect, but biology and history undergraduates were found to give significantly fewer erroneous responses to questions belonging to their respective domains (Cantor & Marsh, 2017). As such, someone who adores Star Trek is much more likely to notice the error in the example given in the introduction of the current article. In contrast, people who are less skilled in reading are even more susceptible to the illusion effect (cf. Hannon & Daneman, 2004). Apart from such interindividual differences, focalisation of the substituted term also seems to play a role (Bredart & Modolo, 1988). Anomalies are more prone to detection when they are part of the focal assertion than when they are part of the presupposition. Furthermore, manipulations of cognitive fluency also seem to alter the probability that an illusion will occur. This can be explained by the notion that fluency influences processing style. As indicated, people tend to rely on effortless means of processing, for reasons of cognitive efficiency and because it generally seems to lead to a correct response. This is also exemplified in the decision-making domain by confirming the common use of fast and frugal heuristics (e.g., Goldstein & Gigerenzer, 2002), which are mental shortcuts that often lead to satisfying solutions. Situations that are fluent, such as encountering seemingly simple questions, may therefore be more likely to yield a Moses illusion. Conversely, low fluency situations engage more systematic processes, making participants rely less on automatic processes and spontaneous intuition (see also Alter, 2013; Alter, Oppenheimer, Epley & Eyre, 2007). As such, less fluent circumstances, such as responding to questions printed in a hard-to-read font, are associated with better detection of distorted questions. Nevertheless, they also appear related to impaired performance on well-formed questions (Song & Schwarz, 2008). Possibly, subjects may be inclined to overanalyse these questions and second-guess their initial reply. As such, subjects will be less susceptible to deception by illusionary questions, but will additionally be inclined to overanalyse well-formed questions.

Interestingly, observations of cognitive disfluency leading to improved performance have also been made in research into FL use. Similar to manipulations of font readability, the use of a non-native language has been found to affect the ease with which information is processed. Indeed, even for non-anomalous text, readers in an FL demonstrate longer sentence reading times, more and longer fixations, shorter saccades, and less word skipping in an FL compared to the NL, suggesting processing disfluency (Cop, Drieghe & Duyck, 2015). This makes adult FL reading comparable to the eye movement patterns found for unskilled NL readers, such as children (Blythe, Liversedge, Joseph, White & Rayner, 2009; Blythe, Häikiö, Bertam, Liversedge & Hyönä, 2011; Rayner, 1986). In fact, for low-proficient bilinguals, reading in an FL is associated with reduced involvement of automatic processes (Favreau & Segalowitz, 1983) and therefore reduced cognitive fluency. In an attempt to investigate heuristic reasoning in FL decision-making, Costa and colleagues (2014) hypothesised that FL processing would either elicit cognitive disfluency or cognitive load. FL-induced cognitive disfluency was expected to reduce heuristics in the same way that incidental experiences of difficulty or disfluency, such as hard-to-read-fonts or having to furrow one's brow, seem to do (e.g., Alter et al., 2007). In contrast, FL-induced cognitive load would endorse more heuristic reasoning and increase bias, as FL processing creates such difficulty, up to a point where the analytic system is overtaxed and individuals revert to employing mental shortcuts. In the end, Costa, Foucart, Arnon, Aparici, and Apesteguia (2014) found that FL use made people less susceptible to cognitive bias, favouring the cognitive fluency hypothesis (see also Keysar, Hayakawa & An, 2012; Winskel, Ratitamkul, Brambley, Nagarachinda & Tiencharoen, 2016).

Given the supposed cognitive nature of the FL effect, the question arises whether it can be generalised beyond the decision-making domain and whether it also applies to anomaly detection. The influence of language on the detection of illusionary questions has briefly been touched upon in the context of a more comprehensive study on the effects of language on moral judgment (Geipel, Hadjichristidis, & Surian, 2015), and results showed that FL use did not improve detection of illusionary questions. However, participants were only presented with one distorted question (i.e., the original Moses illusion) and one control question. Although indicative, the limited number of items does not allow inferences about the effect of language on the detection of semantic anomalies in general. Hence, the aim of the current study was to systematically assess the effect of NL and FL use on the detection of anomalous questions. If FL processing indeed prompts cognitive disfluency, it is expected that participants will be better at detecting semantic anomalies when questions are presented in their FL relative to their NL. Conversely, if FL use should induce cognitive load rather than disfluency, the likelihood that anomalies are detected decreases, similar to less-skilled NL readers being more susceptible to failures of detection (cf. Hannon & Daneman, 2004). Given the limited capacity of controlled processes, high cognitive load as a result of FL use will then cause interference with deeper processing levels, leading to fewer detections.

In order to ascertain whether any processing differences occur online, which cannot be detected via behavioural measures, the current study also opted to include eye movement measures. In general, we suppose there will be no differences for early measures, such as first fixation duration, between detected anomalies and non-detected anomalies and controls; as, during the first pass, participants are not yet aware that the rest of the sentence is inconsistent with the target word (i.e., the anomaly) (cf. Bohan & Sanford, 2008; Daneman et al., 2007). We do expect a difference for later measures, such as more regressions in (as well as longer) total reading times, as participants tend to look back to the detected anomaly. We also believe that undetected anomalies will be encoded and not skipped (see Reder & Kusbit, 1991). Regarding the effect of language, our assumption is that overall, participants reading in their FL will require longer total reading time and make more and longer fixations for both non-anomalous (cf. Cop et al., 2015) and anomalous sentences, but we do not assume that anomaly detection will interact with language.

https://doi.org/10.1017/S1366728921000882 Published online by Cambridge University Press

Method

Participants

Twenty-one native Dutch-speaking psychology students (M_{age} = 19.30, SD_{age} = 1.71; 16 females) at Ghent University participated in our study in exchange for course credit. Participation was restricted to individuals with Dutch as their native language and English as a foreign language, having obtained formal English instruction from their second year of secondary school onward, as is standard in the Dutch-speaking part of Belgium. To take part, they were also required to score at least 60% on the English version of the LexTALE language proficiency scale (Lemhöfer & Broersma, 2011). This criterion – considered as a lower bound for upper intermediate FL users – was set to ensure that participants had sufficient FL knowledge to understand and answer all questions. One female participant did not meet this criterion and was excluded from the analyses. LexTALE was also completed in NL Dutch and, as predicted, scores on the English version of the LexTALE were generally lower (Dutch: M = 87.25, SD = 7.22; English: M = 81.19, SD = 10.65).

Materials

Stimuli

The set of target questions (see Appendix A) comprised both newly formed questions and adaptations from questions used in previous studies (Barton & Sanford, 1993; Büttner, 2007; Cantor & Marsh, 2017; Cook et al., 2018; Erickson & Mattson, 1981; Hannon & Daneman, 2001; Reder & Kusbit, 1991). There were 44 questions and each of these had four variants (Language x Type): NL non-anomalous, NL anomalous, FL non-anomalous, and FL anomalous. In anomalous questions, one word was substituted by a semantically related word, creating an incorrect sentence. There were five kinds of substitutions; proper nouns (e.g., 'Noah' replaced by 'Moses'), nouns (e.g., 'vowel' replaced by 'consonant'), verbs (e.g., 'see' replaced by 'hear'), adjectives/adverbs (e.g., 'extinct' replaced by 'endangered'), and numerals (e.g., 'three' replaced by 'five'). For example, a non-anomalous sentence in English would have been "What do you call a year in which February has one day *more* than usual?", whereas the anomalous trial with the adverb substituted would state "What do you call a year in which February has one day *less* than usual?".

Translated variants closely resembled each other in syntactic structure and within-language variants only differed in one target word substitution. Target length, target frequency, and sentence length were matched for anomalous and non-anomalous trials (all ps > .301). Word frequencies were obtained from the SUBTLEX-NL corpus (Keuleers, Brysbaert, & New, 2010) for Dutch and from the SUBTLEX-UK corpus (van Heuven, Mandera, Keuleers, & Brysbaert, 2014) for English. Target length and sentence length were also matched across language condition (ps > .280), but there was a difference for target frequency, with FL targets more frequently occurring than NL targets ($F_{1,174}$ = 40.73, p < .001). All values are reported in Table 1. There were four versions of stimuli sets, each set containing one of the four variants (counterbalanced) plus 40 filler questions. Participants were randomly assigned to one of the four stimuli sets. All sentences can be consulted via Open Science Framework and are freely available to use in future research.

Table 1. Sentence matching values per anomaly and language condition (standard deviations between parentheses).

	Anomalous		Non-anomalous	
	NL	FL	NL	FL
Target length[1]	7.71 (3.07)	7.14 (2.53)	7.32 (2.51)	7.00 (2.75)
Target frequency[2]	2.27 (0.97)	3.23 (0.94)	2.46 (1.04)	3.38 (0.98)
Sentence length[3]	14.98 (5.51)	15.66 (5.40)	15.05 (5.47)	15.68 (5.39)

[1]Target length is expressed in number of characters; [2]Target frequency was retrieved from the corpora SUBTLEX-NL and SUBTLEX-UK: [3]Sentence length is expressed in number of words.

Apparatus

Eye movements were recorded with an EyeLink 1000 Plus system (SR Research, Canada) with a sampling rate of 1 kHz. Participants were asked to place their head on the forehead and chin rest. The presentation of the questions and recording of the eye movements were implemented using Experiment Builder (SR Research Ltd.). Reading was binocular, but eye movements were recorded from the dominant eye only. Questions were displayed as a whole in the centre of the screen and verbal responses to the questions were registered by the experimenter. Participants only spoke after stimulus presentation (thus not during eye movement recording) and were instructed to move as little as possible.

Eye movements

The measures recorded for target processing were (a) first fixation duration: the length of first fixation within an area of interest; (b) gaze duration: the sum of fixations from the time that an area is first entered from the left to the time that the region is first exited to the right; (c) skipping: the lack of fixation into an IA; (d) total reading time: the total amount of time spent fixating within an area of interest (IA); (e) number of regressions: the total number of regressions into an IA from any region to the right.

Post check

Participants also completed a short paper-and-pencil test in their native language Dutch about the questions in the experiment. The test was provided in multiple-choice format, where participants completed sentences with a missing key word or responded to the question with the non-anomalous term. Three alternatives were always provided: namely, the non-anomalous answer, the substitution (which was also used in the illusionary questions), and another semantically related word. If they were uncertain about the response, they were instructed to leave it open. Participants were told to no longer look for ill-formed sentences.

Procedure

Each participant completed an experimental block in Dutch (NL) and one in English (FL). Participants were randomly assigned to a certain block order. Within one block, there were 11 questions with a substitution (Type: anomalous), 11 questions without a substitution (Type: non-anomalous), and 20 filler questions. Participants were allowed to answer all questions in the language of their choice.

At the start of the experiment, participants were informed that questions were going to be presented on screen, which they would have to answer aloud. Importantly, they were told that some of the questions may be ill-formed and could therefore not be answered.

https://doi.org/10.1017/S1366728921000882 Published online by Cambridge University Press

Ill-formed questions were illustrated by three example questions: one question with a substituted proper noun, one question with a substituted verb, and one question with a substituted adjective. This way, it would be clear that anomalies were not restricted to a specific part of speech. For each example question, an explanation was provided about why the question was ill-formed. Participants were instructed to respond 'WRONG' to such questions. If they believed a question did not contain any substitutions, they were required to give the non-anomalous answer. If they did not know the answer, they could respond with 'DON'T KNOW'. The task was self-paced, but participants were instructed to answer the questions as well and as quickly as possible. Three practice trials were provided in order to familiarise participants with question presentation and responses. No feedback was given either during or after the experiment.

Both experimental blocks were preceded by a 9-point calibration of the eye tracker. Within each block, each trial started with the presentation of the question on screen. When a participant had read the question, he or she pressed the space bar and saw a blank screen. As instructed, they were allowed to further think about the answer and respond to the question when this blank screen was shown. After each response, a drift correction was included so that recalibration could be performed if needed. Before moving on to the second block, participants were allowed to take a short break and move their head from the chinrest (followed by recalibration).

Results

Behavioural analysis

Responses were considered illusions if participants answered an anomalous question with its illusionary response, despite the fact that the subsequent knowledge test confirmed they should have been able to detect the substitution. If participants indicated the incorrect answer in the subsequent knowledge test, the corresponding question was excluded from analysis. Additionally, because unambiguous categorisations could be made solely for trials in which participants explicitly answered the illusionary question with 'WRONG' or the illusionary answer, only these trials were taken into account when calculating the proportion of illusions. To illustrate, for the Moses illusion (i.e., "How many animals did Moses take on the Ark?"), 'WRONG' would be seen as the correct response and 'TWO' as the illusionary (incorrect) response. Another reply, such as 'TEN', would not be taken into account in calculating the proportion of illusions. To reduce the risk of false categorisations, the analysis was restricted to illusions and detections that could unequivocally be derived from participants' responses.

Considering all anomalous trials, uncorrected for the subsequent knowledge test, illusions occurred in 54.55% of all cases, whereas detections were made 32.50% of the time. Participants responded with 'DON'T KNOW' to 10% of the anomalous questions and in 2.95% of the cases, they provided another incorrect or ambiguous response. For the non-anomalous trials, 3.18% were answered incorrectly, 2.50% were erroneously categorised as 'WRONG' (i.e., false detection), and 8.18% were responded to with 'DON'T KNOW'. The remaining 86.14% of the questions were answered non-anomalously.

Observed percentages for each response type can be found in Table 2. After correcting for the subsequent knowledge test (NL = 13.18%, FL = 10.00%) and trials for which the answer was ambiguous (NL = 2.27%, FL = 3.64%) or "DON'T KNOW" (NL = 9.55%, FL = 10.46%), fewer illusions occurred in NL (M = 55.68%, SD = 18.84%) compared to FL (M = 65.09%, SD = 19.39%) when confronted with an anomalous question.

Table 2. Percentages of replies per response type, split for language condition.

	NL	FL	Total
Anomalous questions (uncorrected)			
Illusions	53.18%	55.91%	54.55%
True detections	35.00%	30.00%	32.50%
'Don't know'	9.55%	10.45%	10.00%
Other replies	2.27%	3.64%	2.95%
Anomalous questions (corrected)			
Illusions	55.68%	65.09%	60.39%
True detections	44.32%	34.91%	39.61%
Non-anomalous questions			
Non-anomalous replies	87.27%	85.00%	86.14%
False detections	1.82%	3.18%	2.50%
'Don't know'	8.18%	8.18%	8.18%
Other replies	2.73%	3.64%	3.18%

To analyse whether this difference between language conditions for anomalous sentences was significant, we employed generalised linear mixed-effects regression using the lme4 package (Version 1.1.21, Bates, Maechler, Bolker & Walker, 2015) in R (Version 3.2.2; R Core Team, 2020). The factors Language (two levels: NL and FL) and Substitution Type (five levels: adjective/adverb, noun, numeral, proper noun, and verb) were sum-coded and included as fixed effects along with individual LexTALE English scores as a measure of FL proficiency. Frequency of target words was included in the analysis as a centred covariate. We initially adopted a full random structure (Barr, Levy, Scheepers & Tily, 2013) including random intercepts for participants and items, as well as random slopes for all predictors. In response to convergence issues, however, the model was simplified by stepwise removal of random slopes, eventually leaving only the random intercepts and fixed effects (see Appendix B).

The analysis confirmed that the effect of Language was significant ($\chi^2(1) = 9.7002$, $p = .002$), with a lower probability of an illusion in the NL condition. Furthermore, higher target word Frequency ($\chi^2(1) = 8.3485$, $p = .004$) and higher LexTALE English scores ($\chi^2(1) = 6.3192$, $p = .012$) significantly decreased the probability of an illusion. There was no significant effect of target Substitution Type ($\chi^2(4) = 4.8012$, $p = .308$).

In addition, we ran another mixed-effects model regression on false detections in non-anomalous sentences, with Language as a fixed factor and Participant and Stimulus as random factors. The effect of language was not significant ($\beta = -0.305$, $SE = 0.331$, z-value = −0.923, $p = .356$).

During the course of the analyses, we started contemplating whether illusion type questions had any memorial consequences. Especially when readers fail to detect the anomalies, the error might be considered as truth. To explore whether the illusionary statements had memorial consequences for the subsequent knowledge test (i.e., post check), we looked at the questions

https://doi.org/10.1017/S1366728921000882 Published online by Cambridge University Press

Table 3. Eye tracking descriptives (standard deviations of means provided between parentheses).

Language	Verbal Response	First fixation duration (ms)	Gaze duration (ms)	Skipping rate (%)	Total time (ms)	Regressions (count)
NL	Illusion	278 (177)	732 (788)	0.27	517 (461)	0.447 (0.641)
	Detection	244 (108)	1009 (1489)	0.31	792 (866)	0.603 (0.775)
	Non-anomalous	282 (202)	680 (848)	0.28	508 (550)	0.402 (0.643)
FL	Illusion	308 (225)	794 (1016)	0.27	585 (505)	0.534 (0.802)
	Detection	299 (131)	616 (763)	0.23	827 (712)	0.881 (0.984)
	Non-anomalous	295 (161)	566 (683)	0.21	519 (537)	0.357 (0.572)

incorrectly answered with the substitution term. This was the case for 8.86%, whereas only 2.50% of the questions were incorrectly answered with the other alternative. A binomial test showed this difference was significant (B (11, 0.0886), $p < .001$). Comparing NL (substitution term errors = 11.36%, alternative errors = 1.82%) and FL (substitution term errors = 6.36%, alternative errors = 3.18%), the ratio of questions erroneously answered with the substitution term and those erroneously answered with the other alternative did not differ between languages ($\chi^2(1) = 2.710$, $p = .100$). Although the overall error percentage of errors was lower in FL, a binomial test showed it did not differ significantly from that in the NL (B (21, 0.0659), $p = .070$).

Eye tracking analysis

In analogy to the behavioural data, trials for which the answer was ambiguous or 'DON'T KNOW' or with incorrect answers on the subsequent knowledge were excluded from analysis. In addition, eye tracking data was cleaned in R by removing fixations with a duration of less than 100 ms, as these are considered too short to represent functional processes. This procedure led to a 2.8% loss of data.

We contrasted the eye tracking data gathered during illusions and detections trials to those gathered during trials where participants provided a regular answer to non-anomalous questions. There were five outcome variables of interest: first fixation duration (in ms), gaze duration (in ms), skipping rate (binary), total reading times (in ms), and number of regressions (count). Observed results for each of the outcome variables are found in Table 3. Visualisations of the fixation durations are found in Figure 1.

All outcome variables were modelled through (generalised) linear mixed-effects regression, using the lme4 package (Version 1.1.21, Bates, Maechler, Bolker & Walker, 2015) in R (Version 3.6.1; R Core Team, 2020). To account for the binary/count nature of two of our dependent variables, skipping rates and number of regressions were modelled using a logit and logarithm link function, respectively. First fixation durations, regression path durations, and total times were log-transformed to normalise their distribution. The factors Language (two levels: NL and FL) and Verbal Response (three levels: detection, illusion, non-anomalous question) were sum-coded and included as fixed effects, together with their interaction. Target words' frequency was included as a centred covariate. For all models, we initially adopted a full random structure (Barr et al., 2013) with random intercepts for participants and items, as well as random slopes for Language, Verbal Response and their interaction. If the model failed to converge, the model was simplified by stepwise removal of random slopes. This led to substantially simplified random effect structures, with by-participant and by-item random intercepts for all models, and an additional by-participant random slope for Language for first fixation duration and gaze duration (see Appendix B).

First fixation duration

The model revealed a significant main effect of Language ($F(1, 48.72) = 11.929$, $p = 0.001$) with longer durations in FL. There were, however, no significant effects of Verbal Response ($F(2, 609.82) = 0.028$, $p = .973$) or Frequency ($F(1, 133.58) = 2.552$, $p = .113$), nor a significant interaction between Language and Verbal Response ($F(2, 582.69) = 1.358$, $p = .258$).

Gaze duration

The effects of Verbal Response ($F(1, 586.38) = 3.173$, $p = .043$), Frequency ($F(1, 401.72) = 5.254$, $p = .022$), and the interaction between Language and Verbal Response ($F(2, 571.34) = 3.642$, $p = .027$) were found to be significant. This is in agreement with visual inspection of the data, revealing longer gaze durations for detections in NL. The main effect of Language ($F(1, 61.68) = 0.063$, $p = .803$) failed to reach significance.

Word skipping

Skipping rates of target words revealed significant effects of Language ($\chi^2(1) = 11.477$, $p < .001$) and Frequency ($\chi^2(1) = 16.313$, $p < .001$), with lower skipping rates for targets in FL and lower frequency targets. Crucially, however, there was no significant effect of Verbal Response ($\chi^2(2) = 1.827$, $p = .401$), nor a significant interaction between Language and Verbal Response ($\chi^2(2) = 1.438$, $p = .487$).

Total time

With respect to total reading times of target words, longer durations were observed for sentences in FL ($F(1, 534.97) = 12.524$, $p < .001$) and with lower frequency targets ($F(1, 116.92) = 22.943$, $p < .001$). There was also a significant effect of Verbal Response ($F(2, 615.08) = 14.254$, $p < .001$). Post-hoc Tukey comparisons showed that during detection trials, sentences received significantly longer total reading times compared to illusions ($z = 3.425$, $p = .001$) and non-anomalous questions ($z = -5.349$, $p < .001$). The difference between illusions and non-anomalous questions, however, was not significant ($z = -1.554$, $p = .262$). The interaction between Language and Verbal response ($F(2, 588.48) = 0.963$, $p = .383$) was again not significant.

Regressions

Most regressions to target words were registered for detection targets ($M = 0.732$, $SD = 0.886$), followed by illusion targets ($M =$

https://doi.org/10.1017/S1366728921000882 Published online by Cambridge University Press

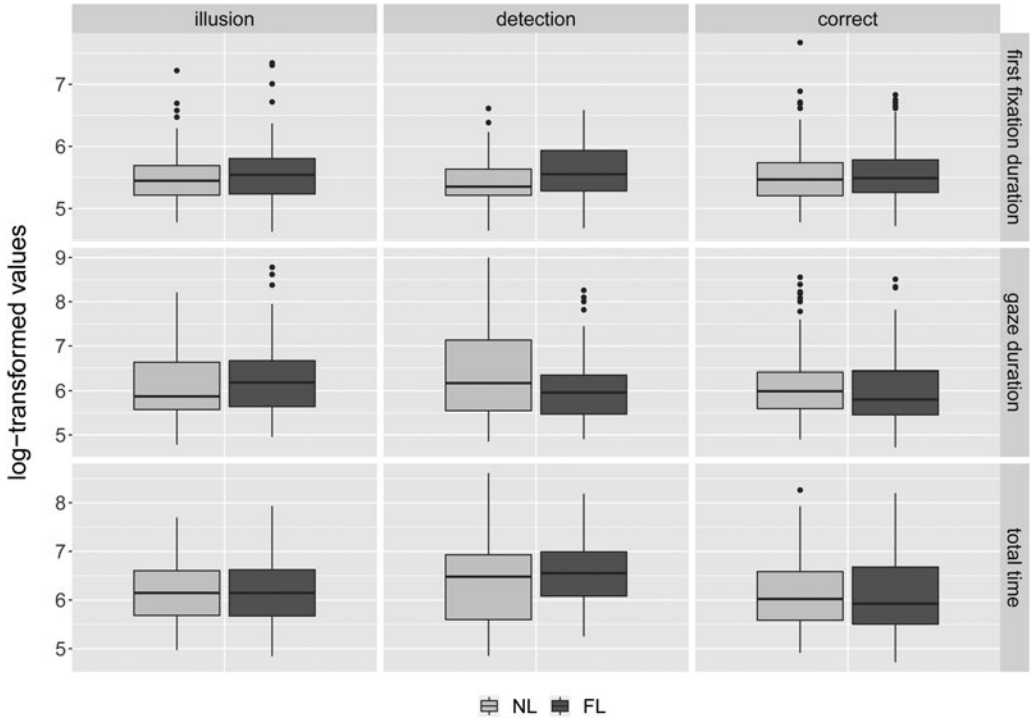


Figure 1. Boxplots of first fixation duration, gaze duration, and total reading time.

0.497, $SD = 0.737$), and non-anomalous targets ($M = 0.380$, $SD = 0.608$). This effect of Verbal Response proved to be significant ($\chi^2(2) = 17.722$, $p < .001$). Tukey comparisons indicated the significant difference was to be found between detections and non-anomalous targets ($z = -4.173$, $p < .001$). There were no differences between detection and illusion trials ($z = 2.008$, $p = .109$), and between illusion and non-anomalous trials ($z = -1.877$, $p = .144$). There were also no main effects of Language ($\chi^2(1) = 2.628$, $p = .105$) or Frequency ($\chi^2(1) = 1.571$, $p = .210$), nor a significant interaction between Language and Verbal Response ($\chi^2(2) = 3.685$, $p = .158$).

Discussion

The aim of the present study was to assess the effect of native (NL) and foreign (FL) language use on the detection of anomalies. Based on hypotheses formulated with regard to decision-making (Costa et al., 2014), we predicted that FL use would either elicit cognitive disfluency leading to increased detection, or induce cognitive load and result in reduced detection. In order to reveal possible differences in anomaly processing, we gathered participants' eye movements as they read the questions in both languages.

The behavioural results of our experiment revealed higher susceptibility to anomalies when questions were formulated in the participants' FL. Contrary to Geipel et al. (2015), who found no differences between FL and NL, we actually observed an FL disadvantage, with higher percentages of illusions. Although this came down to an average of 1.03 illusions in a sample that should be regarded as small, a power analysis confirmed the robustness of the effect (see Appendix C). Our findings are hence not in line with research reporting improved anomaly identification in more cognitively disfluent situations (e.g., Song & Schwarz, 2008). A viable explanation is that employing an FL is more demanding than, for instance, reading in a more illegible font, suggesting that it creates cognitive strain rather than fluency (see Keysar et al., 2012). This strain or load may in turn interfere with deeper processing levels, prompting subjects to engage more in the effortless process of partial semantic matching, making them more prone to disregarding distortions (Reder & Kusbit, 1991). Although not formally tested, the subjective experiences reported by our participants indeed point in this direction; many spontaneously indicated that detecting distortions in FL trials was more difficult relative to NL trials. Relevant to note is that according to the language proficiency test LexTALE (Lemhöfer & Broersma, 2011), our participants can be considered advanced FL users with a mean score of 81%, which corresponds to a C1 level in the Common European Framework of Reference for Languages (Council of Europe, 2001). It is therefore unlikely that the observed language difference can be attributed to a lack of understanding in the FL. Still, we did observe that higher LexTALE English scores significantly decreased the probability of an illusion occurring. Thus, presuming that differences in cognitive load decrease as people are more proficient in their FL, even greater effects of language are to be expected in participants who are less proficient.

In addition to a decrease in detection of distortions, use of an FL also seemed to double the rate of false detections in non-anomalous sentences, although the difference was not significant. This is of course not all that surprising, since the number of false detections was low in general (2.50%). Still, Song and Schwarz (2008) did demonstrate that their cognitively disfluent condition also led to an increase in false detections. The authors related their findings to text familiarity; low familiarity triggers more systematic processing, but also impairs reliance on spontaneous association. In other words, low processing fluency improves performance when one's spontaneous reply is wrong, but impairs

https://doi.org/10.1017/S1366728921000882 Published online by Cambridge University Press

performance when one's spontaneous reply is non-anomalous. Whether the latter is also the case for processing non-anomalous sentences in an FL is still a matter of uncertainty. Truly, increased unfamiliarity induced by FL use does not lead to more methodical processing of distorted material, but there is an indication (although not robust) that it reduces unconstrained processing of non-anomalous material, apparently providing FL readers with the worst of both worlds. This is in stark contrast to the results revealing an FL advantage in decision-making paradigms (e.g., Costa et al., 2014; Keysar et al., 2012).

However, although cognitive fluency is regarded as a possible explanation of reduced engagement in heuristic thinking, the currently prevailing theory is that the FL effect in decision-making stems from an increase in emotional distance (e.g., Hayakawa, Tannenbaum, Costa, Corey & Keysar, 2017). Indeed, the problem statements employed in this strand of research are typically emotion-laden, dealing with the loss of lives and the potential to save lives (e.g., the Asian Disease problem, see Tversky & Kahneman, 1981). On the one hand, lack of emotional attachment may be what causes FL readers to consider the problem in a more systematic manner. On the other hand, and perhaps more likely, the two FL effects may have different underlying mechanisms altogether, with decision-making outcomes being entirely related to an emotional component of cognition and the current FL disadvantage reflecting a purely cognitive issue, such as impartial semantic matching due to a reduction in cognitive fluency.

Apart from the behavioural differences observed in susceptibility, the eye tracking data provided more insight into the online processing of distortions in both language conditions. For first fixation, analyses revealed longer fixations durations in the FL as opposed to the NL but no differences between detections and illusions. This is in line with previous research suggesting no effects of anomaly detection on early measures (Bohan & Sanford, 2008; Daneman et al., 2007). Other main effects of language included longer total reading times and smaller skipping rates in the FL condition. These all point into the same direction as previous research, indicating that readers indeed have more difficulty processing text written in their FL (e.g., Cop et al., 2015) and may experience cognitive load. Looking at the main effects of target for the later measures, we observed longer gaze durations for detection targets, longer total reading times for sentences with detection targets as opposed to those with illusion and non-anomalous targets, and more regressions to detection targets than to non-anomalous targets. Importantly, there was no effect of target on word skipping, and total reading times and regressions did not differ between illusions and non-anomalous questions, confirming that undetected distortions are indeed read and processed similarly to control words, verifying the theory that the mismatch between input and memory representations does not get noticed (see Reder & Kusbit, 1991). Regarding interactions between language and response, analyses only revealed longer gaze durations for detections in the NL. Apart from this, eye movements in the FL condition did not differ from those in the NL condition in relation to detection, illusion, and non-anomalous trials. This indicates that although FL users are more prone to falling for illusions and are overall less fluent readers, their reading patterns are similar to those of NL users when they are exposed to anomalous questions.

Although not a part of our official hypothesis, we started contemplating during the course of the study about whether our Moses illusion type questions had any memorial consequences. For episodic memory, it has long been established that incorrect suppositions in questions may alter subjects' memories of witnessed events (see Loftus, 1975). Within the domain of general knowledge, exposure to factual errors in story reading (Fazio & Marsh, 2008) and to multiple-choice lures (Marsh, Roediger, Bjork & Bjork, 2007) has been found to affect memory. The reader fails to detect the fallacy in the story or selects the wrong answer on the multiple-choice test, and the error afterwards comes to mind fluently and is considered as truth (Kelley & Lindsay, 1993). And indeed, even for the Moses illusion a similar effect has been ascertained; participants used substituted terms embedded in the anomalous questions to later answer general knowledge questions (Bottoms, Eslick & Marsh, 2010). We observed a similar outcome with the more often answered by the substitution term than by the other alternative in the post check knowledge test. Importantly, there was no effect of language, suggesting that use of an FL does not lead to increased disadvantages in false memory. Notable, though, is that the method of determining illusion percentages based on a post check may actually be underestimating the magnitude of the Moses illusion. Although analysing the data in this manner probably did not influence the effect of our language manipulation, subjects' relative inability to detect anomalies may actually be closer to the percentages that were not adjusted for the post check.

To conclude, the present study demonstrates that in contrast with the advantageous FL effects repeatedly observed within decision-making, semantic distortions in reading are more often overlooked when text is formulated in an FL. In absence of discernably disparate reading patterns between FL and NL, we argue that the observed difference in susceptibility is likely to be attributed to the cognitive load that is imposed by use of a non-native language, which in turn increases the tendency to engage in partial semantic processing. Nevertheless, due to the small sample and the moderate number of trials, follow-up replication research is advisable.

Ethical considerations

All subjects were healthy adults and the techniques employed non-invasive. Participation was entirely voluntarily and accord was obtained through information given by the experimenter and a document of informed consent. This document was read and signed before task administration commenced.

Open practise statement

The materials and data are available at Open Science Framework (URL: https://osf.io/3qs9g/?view_only=3a9a3b34fb8e4b0f8b21bd35ce1c735d; DOI 10.17605/OSF.IO/3QS9G).

Acknowledgements. This study was made possible by LEMMA (Language, Education, and Memory in Multilingualism and Academia), a Concerted Research Actions fund (GOA – BOF13/GOA/032) of Ghent University. S. Dhaene and E. Woumans contributed to study conception and design, as well as the literature search. S. Dhaene and N. Dirix contributed to the eye tracking design, programming of the eye tracking material, and analyses of the eye tracking data. S. Dhaene and H. Van Marcke analysed the behavioural and eye tracking data. E. Woumans drafted the manuscript.

References

Alter AL (2013) The benefits of cognitive disfluency. *Current Directions in Psychological Science*, **22**(6), 437–442. doi:10.1177/0963721413498894

https://doi.org/10.1017/S1366728921000882 Published online by Cambridge University Press

Alter AL, Oppenheimer DM, Epley N and Eyre RN (2007) Overcoming intuition: metacognitive difficulty activates analytic reasoning. *Journal of Experimental Psychology: General*, **136**(4), 569–567. doi:10.1037/0096-3445.136.4.569

Barr DJ, Levy R, Scheepers C and Tily HJ (2013) Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, **68** (3), 255–278. doi:10.1016/j.jml.2012.11.001

Barton SB and Sanford AJ (1993) A case study of anomaly detection: Shallow semantic processing and cohesion establishment. *Memory & cognition*, **21** (4), 477–487.

Bates D, Maechler M, Bolker B and Walker S (2015) *lme4 : Linear Mixed-Effects Models Using Eigen and S4*. R package version 1.1–10, http://CRAN.R-project.org/package=lme4.

Bates D, Maechler M, Bolker B and Walker S (2015) Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, **67** (1), 1–48. doi:10.18637/jss.v067.i01

Blythe HI, Liversedge SP, Joseph HS S. L., White SJ and Rayner K (2009) Visual information capture during fixations in reading for children and adults. *Vision Research*, **49** (12), 1583–1591. doi: 10.1016/j.visres.2009.03. 015 0.3758/BF03197179

Blythe HI, Häikiö T, Bertam R, Liversedge SP and Hyönä J (2011) Reading disappearing text: Why do children refixate words? *Vision Research*, **51** (1), 84–92. doi:10.1016/j.visres.2010.10.003

Bohan J and Sanford A (2008) Semantic anomalies at the borderline of consciousness: An eye-tracking investigation. *The Quarterly Journal of Experimental Psychology*, **61**(2), 232–239. doi:10.1080/17470210701617219

Bottoms HC, Eslick AN and Marsh EJ (2010) Memory and the Moses illusion: Failures to detect contradictions with stored knowledge yield negative memorial consequences. *Memory*, **18**, 670–678.

Braze D, Shankweiler D, Ni W and Palumbo LC (2002) Reader's eye movements distinguish anomalies of form and content. *Journal of Psycholinguistic Research*, **31**, 25–44.

Bredart S and Modolo K (1988) Moses strikes again: Focalization effect on a semantic illusion. *Acta Psychologica*, **67** (2), 135–144. doi:10.1016/0001-6918(88)90009-1

Büttner AC (2007) Questions versus statements: Challenging an assumption about semantic illusions. *The Quarterly Journal of Experimental Psychology*, **60** (6), 779–789. doi:10.1080/17470210701228744

Cantor AD and Marsh EJ (2017) Expertise effects in the Moses illusion: detecting contradictions with stored knowledge. *Memory*, **25** (5), 220–230. doi:10.1080/09658211.2016.1152377

Cipolletti H, McFarlane S and Weissglass C (2016) The moral foreign-language effect. *Philosophical Psychology*, **29** (1), 23–40. doi:10.1080/09515089.2014.993063

Cook AE, Walsh EK, Bills MA, Kircher JC and O'Brien EJ (2018) Validation of semantic illusions independent of anomaly detection: evidence from eye movements. *The Quarterly Journal of Experimental Psychology*, **71** (1), 113–121. doi:10.1080/17470218.2016.1264432

Cop U, Drieghe D and Duyck W (2015) Eye movement patterns in natural reading: A comparison of monolingual and bilingual reading of a novel. *PLoS One*, **10** (8), e0134008.

Costa A, Foucart A, Arnon I, Aparici M and Apesteguia J (2014) "Piensa" twice: On the foreign language effect in decision making. *Cognition*, **130**, 236–254. doi:10.1016/j.cognition.2013.11.010

Council of Europe (2001) Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Council of Europe.

Daneman M, Lennertz T and Hannon B (2007) Shallow semantic processing and cognitive processes. *Language and Cognitive Processes*, **22**, 85–105.

Erickson TD and Mattson ME (1981) From words to meaning: A semantic illusion. *Journal of Verbal Learning and Verbal Behavior*, **20** (5), 540–551. doi:10.1016/S0022-5371(81)90165-1

Favreau M and Segalowitz NS (1983) Automatic and controlled processes in the first- and second-language reading of fluent bilinguals. *Memory & Cognition*, **11** (6), 565–574. doi:10.3758/BF03198281

Fazio LK and Marsh EJ (2008) Slowing presenta-tion speed increases illusions of knowledge. *Psychonomic Bulletin & Review*, **15**, 180–185.

Geipel J, Hadjichristidis C and Surian L (2015) How foreign language shapes moral judgment. *Journal of Experimental Social Psychology*, **59**, 8–17. doi:10.1016/j.jesp.2015.02.001

Goldstein DG and Gigerenzer G (2002) Models of ecological rationality: The recognition heuristic. *Psychological Review*, **109** (1), 75–90.

Hannon B and Daneman M (2001) Susceptibility to semantic illusions: An individual-differences perspective. *Memory & cognition*, **29** (3), 449–461. doi:10.3758/BF03196396

Hannon B and Daneman M (2004) Shallow processing of text: An individual-differences account. *Discourse Processes*, **37**, 187–204

Hayakawa S, Tannenbaum D, Costa A, Corey JD and Keysar B (2017) Thinking more or feeling less? Explaining the foreign-language effect on moral judgement. *Psychological Science*, **10**, 1387–1397. doi:1o0i.1or1g7/170/.01197576/709576719776127079240494 4

Hothorn T, Bretz F and Westfall P (2008) Simultaneous Inference in General Parametric Models. *Biometrical Journal*, **50**(3), 346–363.

Kelley CM and Lindsay DS (1993) Remembering mistaken for knowing: Ease of retrieval as a basis for confidence in answers to general knowledge questions. *Journal of Memory and Language*, **32**, 1–24.

Keuleers E, Brysbaert M and New B (2010) SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles. *Behavior Research Methods*, **42** (3), 643–650. doi:10.3758/BRM.42.3.643

Keysar B, Hayakawa S and An SG (2012) The foreign language effect: Thinking in a foreign tongue reduces decision biases. *Psychological Science*, **23**, 661–668. doi:10.1177/0956797611432178

Kuznetsova A, Brockhoff PB and Christensen RH B (2017) lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, **82**(13), 1–26. doi:10.18637/jss.v082.i13

Lemhöfer K and Broersma M (2011) Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods*, **44**, 325–343. doi:10.3758/s13428-011-0146-0

Loftus EF (1975) Leading questions and the eye- witness report. *Cognitive Psychology*, **7**, 560–572.

Marsh EJ, Roediger HL III, Bjork RA and Bjork EL (2007) The memorial consequences of multiple- choice testing. *Psychonomic Bulletin & Review*, **14**, 194–199.

Marsh EJ and Umanath S (2013) Knowledge neglect: Failures to notice contradictions with stored knowledge. In DN Rapp & J Braasch (eds), *Processing inaccurate information: Theoretical and applied perspectives from cognitive science and the educational sciences*. Cambridge, MA: MIT Press.

Rayner K (1986) Eye movements and the perceptual span in beginning and skilled readers. *J Experimental Child Psychology*, **41**, 211–236.

Reder LM and Kusbit GW (1991) Locus of the Moses Illusion: Imperfect encoding, retrieval, or match? *Journal of Memory and Language*, **30** (4), 385–406. doi:10.1016/0749-596X(91)90013-A

Shafto M and MacKay DG (2000) The Moses, Mega-Moses, and Armstrong illusions: Integrating language comprehension and semantic memory. *Psychological Science*, **11**, 372 378.

Song H and Schwarz N (2008) Fluency and the detection of misleading questions: Low processing fluency attenuates the Moses Illusion. *Social Cognition*, **26** (6), 791–799. doi:10.1521/soco.2008.26.6.791

Tversky A and Kahneman D (1981) The framing of decisions and the psychology of choice. *Science*, **211** (4481), 453–458.

Umanath S, Dolan PO and Marsh EJ (2014) Aging and the Moses Illusion: Older adults fall for Moses but if asked directly, stick with Noah. *Memory*, **22**, 481–492.

van Heuven WJ, Mandera P, Keuleers E and Brysbaert M (2014) SUBTLEX-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology*, **67** (6), 1176–1190. doi:10.1080/17470218.2013.850521

Van Oostendorp H and De Mul S (1990) Moses beats Adam: A semantic relatedness effect on a semantic illusion. *Acta Psychologica*, **74** (1), 35–46. doi:10.1016/0001-6918(90)90033-C

Weiss AF, Kretzschmar F, Schlesewsky M, Bornkessel-Schlesewsky I and Staub A (2017) Comprehension demands modulate re-reading, but not first pass reading behavior. *The Quarterly Journal of Experimental Psychology*, 1–37. doi:10.1080/17470218.2017.1307862

Winskel H, Ratitamkul T, Brambley V, Nagarachinda T and Tiencharoen S (2016) Decision-making and the framing effect in a foreign and native language. *Journal of Cognitive Psychology*, **28**, 427–436. doi:10.1080/20445911.2016.1139583

https://doi.org/10.1017/S1366728921000882 Published online by Cambridge University Press

Appendix A

Target questions

A Dutch non-anomalous
B English non-anomalous
C Dutch anomalous
D English anomalous

1a. Welke uitgestorven diersoort speelt de hoofdrol in de film 'Jurassic Park'?
1b. Which extinct species plays the leading part in the movie 'Jurassic Park'?
1c. Welke bedreigde diersoort speelt de hoofdrol in de film 'Jurassic Park'?
1d. Which endangered species plays the leading part in the movie 'Jurassic park'?

2a. Wat is de volledige naam van de lachende 'Mona', één van Da Vinci's meest bekende schilderijen?
2b. What is the full name of the smiling 'Mona', one of Da Vinci's most famous paintings?
2c. Wat is de volledige naam van de lachende 'Mona', één van Rembrandt's meest bekende schilderijen?
2d. What is the full name of the smiling 'Mona', one of Rembrandt's most famous paintings?

3a. Welke kleur heeft de robot in het logo van Android, het mobiele besturingsysteem van Google?
3b. What colour is the robot in the logo of Android, the mobile operating system by Google?
3c. Welke kleur heeft de robot in het logo van Android, het mobiele besturingsysteem van Windows?
3d. What colour is the robot in the logo of Android, the mobile operating system by Windows?

4a. Welk tactiele schrijfsysteem met verhoogde puntjes op papier werd ontwikkeld zodat blinde mensen kunnen lezen met de vingertoppen?
4b. What tactile writing system with raised dots on paper was developed so that blind people can read with the fingertips?
4c. Welk tactiele schrijfsysteem met verhoogde puntjes op papier werd ontwikkeld zodat dove mensen kunnen lezen met de vingertoppen?
4d. What tactile writing system with raised dots on paper was developed so that deaf people can read with the fingertips?

5a. Welk deel van zijn lichaam heeft de beroemde kunstenaar Vincent Van Gogh afgesneden?
5b. What part of his body has the famous artist Vincent Van Gogh cut off?
5c. Welk deel van zijn lichaam heeft de beroemde kunstenaar Teo Van Gogh afgesneden?
5d. What part of his body has the famous artist Teo Van Gogh cut off?

6a. Welk gele spons leeft volgens het Nickelodeonprogramma in een ananas op de bodem van de zee?
6b. What yellow sponge lives according to the Nickelodeon show in a pineapple at the bottom of the sea?
6c. Welk gele spons leeft volgens het Nickelodeonprogramma in een kokosnoot op de bodem van de zee?
6d. What yellow sponge lives according to the Nickelodeon show in a coconut at the bottom of the sea?
7a. Hoe heet het schoolhoofd van Zweinstein, gekend uit de Harry Potter reeks geschreven door J.K. Rowling?
7b. What is the name of the headmaster of Hogwarts, known from the Harry Potter series written by J. K. Rowling?
7c. Hoe heet het schoolhoofd van Zweinstein, gekend uit de Harry Potter reeks geregiseerd door J.K. Rowling?
7d. What is the name of the headmaster of Hogwarts, known from the Harry Potter series directed by J. K. Rowling?

8a. Hoeveel letters zitten er in het alfabet?
8b. How many letters are there in the alphabet?
8c. Hoeveel nummers zitten er in het alfabet?
8d. How many numbers are there in the alphabet?

9a. In welk bordspel kan je punten scoren door horizontaal en verticaal lettertegels te leggen om woorden te vormen?
9b. In which board game can you score points by horizontally and vertically laying out beige lettertiles to form words?
9c. In welk kaartspel kan je punten scoren door horizontaal en verticaal lettertegels te leggen om woorden te vormen?
9d. In which card game can you score points by horizontally and vertically laying out beige lettertiles to form words?

10a. Wat is de naam van de instelling waar kinderen verblijven wanneer hun ouders gestorven zijn, voordat ze worden afgestaan voor adoptie?
10b. What is the name of the institution where children stay when their parents died, before they are put up for adoption?
10c. Wat is de naam van de instelling waar kinderen verblijven wanneer hun ouders gestorven zijn, voordat ze worden afgestaan voor abortus?
10d. What is the name of the institution where children stay when their parents died, before they are put up for abortion?

11a. Als een vliegtuig met Belgische toeristen crasht op Spaans grondgebied, in welk land zullen de slachtoffers dan waarschijnlijk worden begraven?
11b. If a plane with Belgian tourists crashes on Spanish territory, in which country will the victims most likely be buried?
11c. Als een vliegtuig met Belgische toeristen crasht op Spaans grondgebied, in welk land zullen de overlevenden dan waarschijnlijk worden begraven?
11d. If a plane with Belgian tourists crashes on Spanish territory, in which country will the survivors most likely be buried?

12a. Welke stad is volgens de Romeinse mythologie gesticht door de broers Romulus en Remus?
12b. Which city has according to the Roman mythology been founded by the brothers Romulus and Remus?
12c. Welke stad is volgens de Griekse mythologie gesticht door de broers Romulus en Remus?
12d. Which city has according to the Greek mythology been founded by the brothers Romulus and Remus?

13a. Welk Afrikaanse zoogdier wordt gebruikt als logo voor het chocolademerk Côte d'Or, verwijzend naar de oorsprong van de bonen?
13b. What African mammal is used as a logo for the chocolate brand Côte d'Or, referring to the origin of the beans?
13c. Welk Afrikaanse zoogdier wordt gebruikt als logo voor het chocolademerk Côte d'Or, verwijzend naar de oorsprong van de granen?
13d. What African mammal is used as a logo for the chocolate brand Côte d'Or, referring to the origin of the grains?

14a. Welke democratische politicus schreef geschiedenis als de eerste zwarte president van de Verenigde Staten?
14b. Which democratic politician made history as the first black president of the United States?
14c. Welke republikeinse politicus schreef geschiedenis als de eerste zwarte president van de Verenigde Staten?
14d. Which republican politician made history as the first black president of the United States?

15a. Welke kleur heeft de tekstballon in het WhatsApp logo?
15b. What colour is the speech bubble in the WhatsApp logo?
15c. Welke kleur heeft de envelop in het WhatsApp logo?
15d. What colour is the envelope in the WhatsApp logo?

16a. Hoe heet het kleine stuk rubber aan het einde van een potlood om fouten mee weg te vegen?
16b. What's the name of the little piece of rubber at the end of a pencil to erase mistakes?

https://doi.org/10.1017/S1366728921000882 Published online by Cambridge University Press

16c. Hoe heet het kleine stuk rubber aan het einde van een pen om fouten mee weg te vegen?
16d. What's the name of the little piece of rubber at the end of a pen to erase mistakes?

17a. Madrid is de hoofdstad van welk land?
17b. Madrid is the capital of which country?
17c. Barcelona is de hoofdstad van welk land?
17d. Barcelona is the capital of which country?

18a. Wat is de naam van de Belgische nationale voetbalploeg met spelers als Jan Vertonghen, Toby Alderweireld en Eden Hazard?
18b. What's the name of the Belgian national soccer team with players like Jan Vertonghen, Toby Alderweireld and Eden Hazard?
18c. Wat is de naam van de Belgische nationale voetbalploeg met verdedigers als Jan Vertonghen, Toby Alderweireld en Eden Hazard?
18d. What's the name of the Belgian national soccer team with defenders like Jan Vertonghen, Toby Alderweireld and Eden Hazard?

19a. Wat wordt gevierd met vuurwerk op 31 december?
19b. What is celebrated with fireworks at December 31th?
19c. Wat wordt gevierd met vuurwerk op 31 janurari?
19d. What is celebrated with fireworks at January 31th?

20a. Welke gevaarlijke grijze carnivoren werden in de aandacht gebracht door de film "Jaws"?
20b. What dangerous grey carnivores were brought into the spotlight by the movie "Jaws"?
20c. Welke gevaarlijke grijze herbivoren werden in de aandacht gebracht door de film "Jaws"?
20d. What dangerous grey herbivores were brought into the spotlight by the movie "Jaws"?

21a. Welke klinker mist er in de lijst a, e, i, o, …?
21b. What vowel is missing in the list a, e, i, o, …?
21c. Welke medeklinker mist er in de lijst a, e, i, o, …?
21d. What consonant is missing in the list a, e, i, o, …?

22a. Wat is de naam van de jonge Vlaamse popgroep die onlangs bekroond werd met meerdere MIA's en meest gekend is om hun lied 'Goud'?
22b. What is the name of the young Flemish pop group that has recently been awarded with multiple MIA's and is mostly known for their song 'Goud'?
22c. Wat is de naam van de jonge Vlaamse rockgroep die onlangs bekroond werd met meerdere MIA's en meest gekend is om hun lied 'Goud'?
22d. What is the name of the young Flemish rock group that has recently been awarded with multiple MIA's and is mostly known for their song 'Goud'?

23a. Als een kip 1 ei legt per dag, hoeveel eieren zul je dan hebben na 2 weken?
23b. If a chicken lays 1 egg a day, how many eggs will you have after 2 weeks?
23c. Als een haan elke dag 1 ei legt, hoeveel eieren zul je dan hebben na 2 weken?
23d. If a rooster lays 1 egg a day, how many eggs will you have after 2 weeks?

24a. Welke gele citrusvrucht heeft een zure smaak en wordt vaak gebruikt voor sappen?
24b. Which yellow citrus fruit has a sour taste and is often used for juices?
24c. Welke gele citrusvrucht heeft een bittere smaak en wordt vaak gebruikt voor sappen?
24d. Which yellow citrus fruit has a bitter taste and is often used for juices?

25a. Wat is de naam van de Japanse puzzel waarin de getallen 1 tot en met 9 slechts éénmaal mogen voorkomen, zowel horizontaal als verticaal?
25b. What is the name of the Japanese puzzle in which the numbers 1 to 9 can only occur once, both horizontally and vertically?
25c. Wat is de naam van de Japanse puzzel waarin de getallen 1 tot en met 9 slechts éénmaal mogen voorkomen, zowel horizontaal als diagonaal?
25d. What is the name of the Japanese puzzle in which the numbers 1 to 9 can only occur once, both horizontally and diagonally?

26a. Welke beroemde schip raakte een ijsberg in de Atlantische Oceaan en zonk?
26b. Which famous ship hit an iceberg in the Atlantic Ocean and sank?
26c. Welke beroemde schip raakte een ijsberg in de Pacifische Oceaan en zonk?
26d. Which famous ship hit an iceberg in the Pacific Ocean and sank?

27a. Wie was de vrouwelijke Afro-Amerikaans burgerrechten activiste die weigerde om haar zitplaats in de bus af te staan aan een blanke passagier?
27b. Who was the female African-American civil rights activist who refused to give up her seat in the bus to a white passenger?
27c. Wie was de vrouwelijke Afro-Amerikaans burgerrechten activiste die weigerde om haar zitplaats in de trein af te staan aan een blanke passagier?
27d. Who was the female African-American civil rights activist who refused to give up her seat in the train to a white passenger?

28a. Hoe noem je een jaar waarin februari één dag meer telt dan normaal?
28b. What do you call a year in which February has one day more than usual?
28c. Hoe noem je een jaar waarin februari één dag minder telt dan normaal?
28d. What do you call a year in which February has one day less than usual?

29a. E = mc2 werd voorgesteld door welke Duitse wetenschapper?
29b. E = mc2 was proposed by which German scientist?
29c. E = mc2 werd voorgesteld door welke Poolse wetenschapper?
29d. E = mc2 was proposed by which Polish scientist?

30a. Waarom verandert een kameleon van kleur in de aanwezigheid van een bedreiging?
30b. Why does a chameleon change colour in the presence of a threath?
30c. Waarom verandert een salamander van kleur in de aanwezigheid van een bedreiging?
30d. Why does a salamander change colour in the presence of a threath?

31a. Hoeveel dieren van elke soort nam Noah mee op zijn ark?
31b. How many animals of each kind did Noah take on his ark?
31c. Hoeveel dieren van elke soort nam Mozes mee op zijn ark?
31d. How many animals of each kind did Moses take on his ark?

32a. Voor hoeveel jaar sliep Doornroosje nadat ze haar vinger prikte?
32b. For how many years did Sleeping Beauty sleep after she pricked her finger?
32c. Voor hoeveel jaar sliep Sneeuwwitje nadat ze haar vinger prikte?
32d. For how many years did Snow White sleep after she pricked her finger?

33a. Wat is de naam van de man in het rode pak en lange witte baard die kerstcadeaus uitdeelt vanuit zijn slee?
33b. What is the name of the man in the red suit and long white beard who gives out christmas presents from his sleigh?
33c. Wat is de naam van de man in het rode pak en lange witte baard die verjaardagscadeaus uitdeelt vanuit zijn slee?
33d. What is the name of the man in the red suit and long white beard who gives out birthday presents from his sleigh?

34a. Wat is de naam van de bekende prijs uitgereikt door Zweden voor opmerkelijke bijdragen aan de wetenschap?
34b. What is the name of the famous prize issued by Sweden for remarkable contributions to science?
34c. Wat is de naam van de bekende prijs uitgereikt door Denemarken voor opmerkelijke bijdragen aan de wetenschap?
34d. What is the name of the famous prize issued by Denmark for remarkable contributions to science?

35a. Welke Amerikaanse astronaut was de eerste om te wandelen op de maan?
35b. Which American astronaut was the first to walk on the moon?

https://doi.org/10.1017/S1366728921000882 Published online by Cambridge University Press

35c. Welke Amerikaanse astronaut was de eerste om te wandelen op de zon?
35d. Which American astronaut was the first man to walk on the sun?

36a. Welk groot bruin dier met twee bulten wordt gebruikt om goederen te transporteren door de woestijn?
36b. What large brown animal with two humps is used for transporting goods across the desert?
36c. Welk groot bruin dier met drie bulten wordt gebruikt om goederen te transporteren door de woestijn?
36d. What large brown animal with three humps is used for transporting goods across the desert?

37a. Welke sport maakt gebruik van een zware bal met drie gaten om witte kegels omver te gooien?
37b. What sport makes use of a heavy ball with three holes for knocking down white pins?
37c. Welke sport maakt gebruik van een zware bal met vijf gaten om witte kegels omver te gooien?
37d. What sport makes use of a heavy ball with five holes for knocking down white pins?

38a. Welk continent omvat Argentinië, Bolivië, Peru en Brazilië?
38b. What continent includes Argentina, Bolivia, Peru and Brazil?
38c. Welk land omvat Argentinië, Bolivië, Peru en Brazilië?
38d. What country includes Argentina, Bolivia, Peru and Brazil?

39a. Welke bekende "Marilyn" bleek een affaire te hebben met president Kennedy?
39b. Which well-known "Marilyn" was found to have an affair with president Kennedy?
39c. Welke bekende "Marilyn" bleek een affaire te hebben met premier Kennedy?
39d. What well-known "Marilyn" was found to have an affair with premier Kennedy?

40a. Waarom moet je nooit schuilen onder een boom wanneer je bliksem ziet tijdens een storm?
40b. Why should you never seek shelter under a tree when you see lightning during a storm?
40c. Waarom moet je nooit schuilen onder een boom wanneer je bliksem hoort tijdens een storm?
40d. Why should you never seek shelter under a tree when you hear lightning during a storm?

41a. Welke grote witte beer is te vinden op de noordpool en is één van de voornaamste slachtoffers van de opwarming van de aarde?
41b. What large white bear can be found at the north pole and is one of the main victims of global warming?
41c. Welke grote witte beer is te vinden op de zuidpool en is één van de voornaamste slachtoffers van de opwarming van de aarde?
41d. What large white bear can be found at the south pole and is one of the main victims of global warming?

42a. Hoe heet de president van Frankrijk?
42b. What's the name of the president of France?
42c. Hoe heet de koning van Frankrijk?
42d. What's the name of the king of France?

43a. Hoe wordt het genoemd in honkbal wanneer de speler de bal zo ver slaat met zijn knuppel dat hij tijd heeft om helemaal rond het speelveld te lopen voor de bal is teruggebracht?
43b. What's it called in baseball when the player hits the ball with his bat so far that he has time to run all the way around the playing field before it is returned?
43c. Hoe wordt het genoemd in honkbal wanneer de speler de bal zo ver slaat met zijn racket dat hij tijd heeft om helemaal rond het speelveld te lopen voor de bal is teruggebracht?
43d. What's it called in baseball when the player hits the ball with his racket so far that he has time to run all the way around the playing field before it is returned?

44a. Waarvoor staat het acroniem YOLO, een populair gezegde vaak gebruikt door jongeren?
44b. What does the acronym YOLO stand for, a popular saying often used by teenagers?
44c. Waarvoor staat het antoniem YOLO, een populair gezegde vaak gebruikt door jongeren?
44d. What does the antonym YOLO stand for, a popular saying often used by teenagers?

Filler questions

A Dutch filler question
B English filler question

1a. Welk fictief figuur werd gecreërd door Gepetto de houtsnijder?
1b. Which fictional character was created by Gepetto the woodcarver?

2a. Welk merk van gezoete hazelnootpasta wordt gemaakt door het Italiaanse bedrijf Ferrero?
2b. What brand of sweetened hazelnut spread is manufactured by the Italian company Ferrero?

3a. In welke film is Ariel bevriend met een krab genaamd Sebastiaan?
3b. In which movie is Ariel friends with a crab named Sebastian?

4a. Welke autofabricant maakt de Fiesta, Ka en Focus modellen?
4b. Which car maker makes the Fiesta, Ka and Focus models?

5a. Wat is de naam van Shrek's vrouw in de film "Shrek"?
5b. What is the name of Shrek's wife in the movie "Shrek"?

6a. Welk Schotse meer staat erom bekend een monster te bevatten?
6b. Which Scottish loch is reputed to contain a monster?

7a. Wat is de hoofdstad van België?
7b. What is the capital of Belgium?

8a. Welke pizza wordt belegd met ham en ananas?
8b. What pizza is topped with ham and pineapple?

9a. Welke twee kleuren vormen de vlag van Spanje?
9b. What two colours make up the flag of Spain?

10a. Welke richting gaat tegen de klok in, links of rechts?
10b. Which way is anti-clockwise, left or right?

11a. Welke mannelijke jonge zanger is bekend om liedjes als 'Baby' en 'Love Yourself'?
11b. Which male young singer is known for songs like 'Baby' and 'Love Yourself'?

12a. In welke animatiefilm kan je het lied 'Hakuna Matata' horen?
12b. In which animation movie can you hear the song 'Hakuna Matata'?

13a. Welke kleur bekom je wanneer je rood en geel mengt?
13b. Which color do you get if you mix red and yellow?

14a. Welke witte vloeistof wordt geproduceerd door vrouwelijke dieren om hun jongen te voederen?
14b. What white liquid is produced by female animals to feed their youngs?

15a. Hoe noem je iemand die geen vlees eet?

https://doi.org/10.1017/S1366728921000882 Published online by Cambridge University Press

15b. How do you call someone who doesn't eat meat?

16a. Wat is de naam van de huidige koning van Belgie, getrouwd met koningin Mathilde en vader van prinses Elizabeth?
16b. What's the name of the current king of Belgium, married to queen Mathilde and father of princess Elizabeth?

17a. Welke Belgische superheld is stiekem verliefd op Mega Toby?
17b. What Belgian super hero is secretly in love with Mega Toby?

18a. In welke sport was Muhammad Ali de wereldkampioen?
18b. In which sport was Muhammad Ali the world champion?

19a. Wat is de naam van de voormalige Belgische eerste minister, gekend om zijn rode vlinderdas?
19b. What's the name of the former Belgian prime minister, known for his red bow tie?

20a. Welke zoete, kleverige vloeistof wordt gemaakt door bijen?
20b. What sweet, sticky fluid is made by bees?

21a. Wat is de naam van de cartoonist die Mickey Mouse, Donald Duck en Goofy heeft gecreeërd?
21b. What is the name of the cartoonist who created Mickey Mouse, Donald Duck and Goofy?

22a. Hoeveel zijden heeft een driehoek?
22b. How many sides does a triangle have?

23a. Wat is het Franse woord voor 'dinsdag'?
23b. What is the French word for 'Tuesday'?

24a. Citroenen en limoenen zijn een uitstekende bron van welke vitamine?
24b. Lemons and limes are an excellent source of which vitamin?

25a. Welke letter bevindt zich tussen de letters A en E op een azerty toetsenbord?
25b. Which letter is located between the letters A and E on an azerty keyboard?
26a. Wat is het symbool van de olympische spelen?
26b. What's the symbol of the Olympics?

27a. Welke boze tovenaar jaagt de kleine blauwe Smurfen uit hun dorp?
27b. What evil wizard chases the tiny blue Smurfs out of their village?

28a. Welk land heeft een volkslied getiteld 'Wilhelmus'?
28b. Which country has a national anthem entitled 'Wilhelmus'?

29a. Wat is de naam van het deel van het menselijk skelet dat de hersenen beschermt?
29b. What is the name of the part of the human skeleton which protects the brain?

30a. Wat is de chemische formule van water?
30b. What is the chemical formula of water?

31a. Wat is de voornaam van de Amerikaanse President?
31b. What's the first name of the American President?

32a. Welke sport kan worden geassocieerd met Michael Phelps?
32b. Which sport can be associated with Michael Phelps?

33a. Wat is de naam van de beroemde muur in Duitsland die werd neergehaald in 1989?
33b. What is the name of the famous wall in Germany that was torn down in 1989?

34a. Wat is de naam van het orgaan dat bloed door het lichaam pompt?
34b. What is the name of the organ that pumps blood around the body?

35a. Welk groot standbeeld bevindt zich in New York Harbor en symboliseert vrijheid?
35b. What large statue is located in New York Harbor and symbolizes freedom?

36a. Hoeveel wereldoorlogen waren er in de twintigste eeuw?
36b. How many world wars were there in the twentieth century?

37a. Welke planeer staat het dichtste bij de zon?
37b. Which planet is closest to the sun?

38a. Flamenco, mambo, tango en rumba zijn allen soorten van wat?
38b. Flamenco, mambo, tango and rumba are all types of what?

39a. Wat is de familienaam van VTM nieuwslezer 'Dany'?
39b. What's the last name of VTM news anchor 'Dany'?

40a. Welk metaal is vloeibaar bij kamertemperatuur?
40b. What metal is liquid at room temperature?

Appendix B

Table #. Model results table for the behavioral analysis on anomalous sentences.

	χ^2	*Df*	*p*	
Fixed effects				
Intercept	5.0066	1	0.025250	*
Language	9.7002	1	0.001842	**
Frequency	8.3485	1	0.003860	**
LexTale	6.3192	1	0.011944	*
Substitution Type	4.8012	4	0.308311	
	Variance	*SD*		
Random effects				
Participant				
(Intercept)	0.3893	0.6239		
Item				
(Intercept)	1.8907	1.3750		

https://doi.org/10.1017/S1366728921000882 Published online by Cambridge University Press

Table #. Model results table for all eye-tracking analyses.

	F	*Df*	*Df.res*	*p*	
First fixation duration					
Fixed effects					
Intercept	24462.5976	1	43.12	< 2.2e-16	***
Language	11.9288	1	48.72	0.001154	**
Verbal response	0.0277	2	609.82	0.972697	
Frequency	2.5521	1	133.58	0.112507	
Language x Verbal response	1.3578	2	582.69	0.258045	
	Variance SD				
Random effects					
Participant					
(Intercept)	0.006293	0.07933			
Language	0.000513	0.02265			
Item					
(Intercept)	0.027080	0.16456			
Residual					
	0.157641	0.39704			
	F	*Df*	*Df.res*	*p*	
Gaze duration					
Fixed effects					
Intercept	4378.6816	1	52.40	< 2e-16	***
Language	0.0630	1	61.68	0.80269	
Verbal response	3.1733	2	586.38	0.04259	*
Frequency	5.2543	1	401.72	0.02241	*
Language x Verbal response	3.6418	2	571.34	0.02681	*
	Variance SD				
Random effects					
Participant					
(Intercept)	0.028305	0.16824			
Language	0.002302	0.04798			
Item					
(Intercept)	0.280638	0.52975			
Residual					
	0.348641	0.59046			
	χ^2	*Df*		*p*	
Word skipping					
Fixed effects					
Intercept	45.5542	1		1.485e-11	***
Language	11.4772	1		0.0007045	***
Verbal response	1.8274	2		0.4010308	
Frequency	16.3125	1		5.371e-05	***

https://doi.org/10.1017/S1366728921000882 Published online by Cambridge University Press

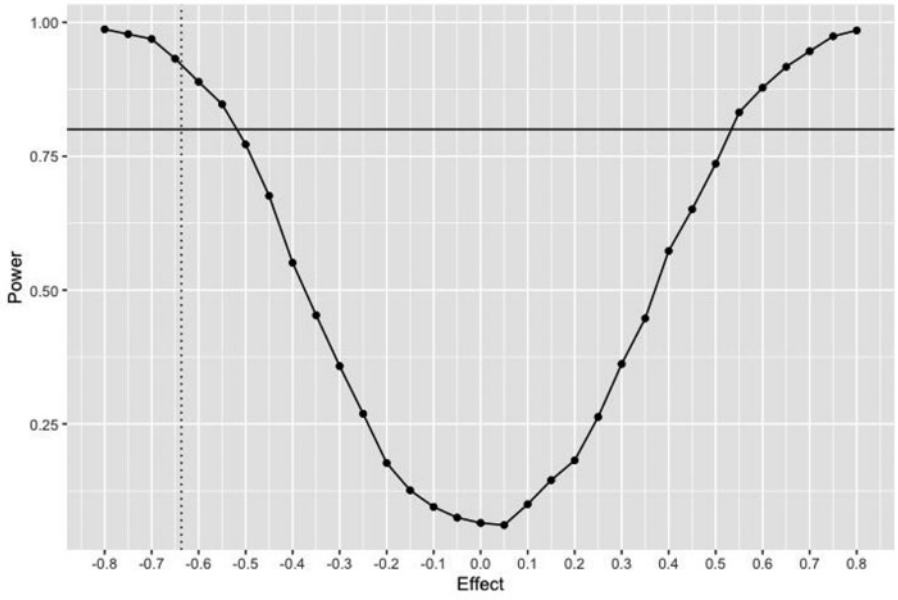
Language x Verbal response	1.4376	2		0.4873442	
	Variance SD				
Random effects					
Participant					
(Intercept)	0.1180	0.3436			
Item					
(Intercept)	0.7403	0.8604			
	F	*Df*	*Df.res*	*p*	
Total time					
Fixed effects					
Intercept	12516.2108	1	39.29	< 2.2e-16	***
Language	12.5241	1	534.97	0.0004367	***
Verbal response	14.2538	2	615.08	8.888e-07	***
Frequency	22.9425	1	116.92	4.940e-06	***
Language x Verbal response	0.9626	2	588.48	0.3824969	
	Variance SD				
Random effects					
Participant					
(Intercept)	0.02188	0.1479			
Item					
(Intercept)	0.05344	0.2312			
Residual					
	0.39667	0.6298			
	χ^2	*Df*		*p*	
Number of regressions					
Fixed effects					
Intercept	42.6848	1		6.431e-11	***
Language	2.6280	1		0.1049924	
Verbal response	17.7221	2		0.0001418	***
Frequency	1.5708	1		0.2100973	
Language x Verbal response	3.6852	2		0.1584071	
	Variance SD				
Random effects					
Participant					
(Intercept)	0.01201	0.1096			
Item					
(Intercept)	0.63896	0.7994			

https://doi.org/10.1017/S1366728921000882 Published online by Cambridge University Press

Appendix C

Post hoc power analyses for the mixed effects model reported in the paper for the behavioural data was based on the information provided by Brysbaert and Stevens, 2018. Power simulations were conducted via the *simr* package in R. In *simr*, power is calculated by repeating the following three steps: (i) simulate new values for the response variable using the model provided; (ii) refit the model to the simulated response; (iii) apply a statistical test to the simulated fit. In this setup, the tested effect is known to exist, so every positive test is a true positive and every negative test is a Type II error. Taking computer power and time constraints into account, simulations were done with 1000 iterations per effect size. As fitted effect of Language (which would be considered as a population value in the simulations) was equal to -0.6369, the effect was varied from null effect to +/-0.8 by steps of 0.05. This enabled us to determine how large the language effect would have to be in order to detect it. This post hoc power analysis (setting all other population values to the effect sizes that were observed) confirmed that the current design was sufficiently powered to detect the observed language effect (see plot below).

Fixed effects fitted model	Estimate	SE	z-value	p-value
Intercept	1.1080	0.4952	2.238	0.025
Language	−0.6369	0.2045	−3.115	0.002
Frequency	−0.8934	0.3092	−2.889	0.004
LexTale	−4.9114	1.9538	−2.514	0.012
Substitution Type 1	0.7307	0.7001	1.044	0.297
Substitution Type 2	−0.7414	0.5591	−1.326	0.185
Substitution Type 3	−0.4371	1.0482	−0.417	0.677
Substitution Type 4	−0.7123	0.6316	−1.128	0.259



https://doi.org/10.1017/S1366728921000882 Published online by Cambridge University Press