Language and Cognition (2024), 1–25
doi:10.1017/langcog.2024.51

ARTICLE

The how and the when of semantic illusions in native and non-native languages

Ana I. Fernandes[1], Juliana Novo Gomes[2], Juan Haro[3], Cristina Flores[4] and Montserrat Comesaña[1]

[1]Psycholinguistics Research Line, CIPsi, School of Psychology, University of Minho, Braga, Portugal; [2]Center of Linguistics, University of Porto, Porto, Portugal; [3]Research Center for Behavior Assessment (CRAMC) and Department of Psychology, Universitat Rovira i Virgili, Tarragona, Spain and [4]Center for Humanities, School of Arts and Humanities, University of Minho, Braga, Portugal
Corresponding author: Ana I. Fernandes; Email: pg50190@alunos.uminho.pt

(Received 25 May 2023; Revised 18 June 2024; Accepted 27 June 2024)

Abstract

The nature and processing of semantic illusions (SI; when speakers fail to notice an anomalous word in a sentence that is contextually perfectly aligned with world knowledge) have been largely studied during first language comprehension. Although this issue is not free of controversy, findings sustain The Node Structure Theory, according to which SI is a phonological and/or semantic priming effect which occurs due to phonological and/or semantic links existing between the correct and the anomalous word. However, the question as to whether the same underlying mechanisms can be found in bilinguals and whether the effect is modulated by age of language acquisition (AoA) and language dominance remains unexplored. The aim of this study was to examine this issue on sequential European Portuguese-German bilinguals (and their respective control groups) using a self-paced reading paradigm. The sentences' language, AoA (early vs. late), and type of target word used (correct vs. anomalous) were manipulated. Results showed the occurrence of SI, independently of language and AoA. Therefore, findings suggest that SI occur due to a semantic overlap between critical words and are similarly processed in L1 and L2.

Keywords: bilingualism; language processing; Rapid Serial Visual Presentation; self-paced paradigm; semantic illusions

1. Introduction

When confronted with the question 'How many animals of each kind did Moses take on the Ark?' most people tend to answer 'two' and do not notice that it was *Noah* and not *Moses* who took the animals on the Ark. This phenomenon is known as the Moses-Illusion, a sort of semantic illusion in which readers often fail to notice the anomalous word (*Noah*) in a sentence that is perfectly aligned with world knowledge (Erickson & Mattson, 1981). At first, the sentence seems to be semantically correct, but upon closer examination, a semantic anomaly can be detected. The representational nature and processing of semantic illusions have been extensively studied during first language (L1) comprehension (Gomes & Maia, 2024), as we will see later on. However, the question regarding the nature of semantic illusions that may arise during second language (L2) processing remains almost unexplored (Dhaene et al., 2022; Vaessen, 2017). Additionally, whether the age of L2 acquisition and language dominance modulate the effect or not requires further investigation. The present research aimed to examine this issue in the two languages spoken by early and late European Portuguese (EP; L1)-German (GER; L2) bilinguals with different degrees of L2 proficiency, and thus varying language dominance (referring to the degree of balance between the two languages of the bilingual speaker). Two groups of native speakers of EP (EP speakers with no knowledge of GER) and of GER (GER speakers with no knowledge of EP) were also examined as control groups.

1.1. Semantic illusions in first language processing

In order to understand how semantic illusions emerge, it is crucial to discuss existing theories and influencing factors on the occurrence of semantic illusions in an L1.

In one of the first studies on the so-called 'Moses-Illusion', Erickson and Mattson (1981) contradicted general belief showing that participants do not process the critical name in a semantic illusion, neither at the phonemic nor at the graphemic levels, and that they respond to the experimental question without acknowledging the illusive name. In their first experiment, participants were asked to read Moses-Illusion-like sentences out loud (e.g., *How many animals of each kind did Moses take on the Ark?)* and to answer aloud questions as fast as possible to ensure that the incongruous word was encoded at least at the phonemic level. Since semantic illusions were observed, even though sentences were read out loud, the authors argued that the inconsistent name (e.g., '*Moses*') was indeed encoded during sentence processing. Hence, the authors proposed that participants' failure to recognize inaccuracies during the processing of Moses-like illusions stems from the focus of the questions (e.g., 'How many animals' *vs.* 'Who took the animals'). Therefore, the second experiment was conducted to test this hypothesis. They shifted the focus of the questions away from the inconsistent name to examine whether the illusion only arose in a particular form of questions. Target questions used in the first experiment were turned into statements, such as '*Moses took two animals of each kind on the Ark*'. Participants had to read statements and circle 'true', 'false' or 'I don't know' as response in a booklet. Although focalization attenuated the percentage of semantic illusions, illusions were still observed.

Additionally, Erickson and Mattson (1981) stated that the Moses-Illusion seems to work in specific combinations only, since the anomaly is easily identified by participants whenever the incongruous word is not related to the correct word (e.g., '*How many animals of each kind did George Washington take on the Ark*?'). Frequently, inconsistent names used in an illusion context are based on either phonological or semantic similarity with the correct word, resulting in two possible hypotheses. Thinking about the Moses-Illusion, *Moses* and *Noah* contain two syllables with the first syllable being stressed in both words and having an 'o' sound at the beginning of the word. In this case, the illusion could be caused by phonological similarity. On the other hand, *Moses* and *Noah* share semantic features; both are biblical figures who received messages from God and are linked to the sea. Therefore, the Moses-Illusion could also be due to semantic similarities. The procedure of the third experiment was similar to the second one – different variations of the questions were presented by inserting other names. In the Moses-Illusion, for example, the phonological similarity hypothesis was tested by replacing the illusive name '*Moses*' with '*Adam*' or '*Abraham*'. The three names were chosen because they vary in the degree of phonological similarity and because they are similar at the semantic level, representing biblical names that appear in the Old Testament. Likewise, to test the semantic similarity hypothesis, '*Noah*' was replaced by a nonbiblical figure, namely '*Nixon*'. Results indicated that a semantic illusion is more likely to occur if inconsistent and correct target words have a semantic resemblance. No significant differences between the three levels of phonological similarity (for example, *Moses*, *Adam* and *Abraham*) were found. The authors concluded that the Moses-Illusion seems to occur when a bunch of semantic features are connected to the content of the sentence, confirming the semantic similarity hypothesis. Thus, when a participant reads something about an Ark, a semantic network connected to the concept of the Ark will be activated, including biblical figures, such as *Moses*, *Adam* or *Abraham*.

The seminal study by Erickson and Mattson (1981) led many studies to (re) examine the semantic similarity hypothesis (e.g., Reder & Kusbit, 1991; Van Jaarsveld et al., 1997; Van Oostendorp & de Mul, 1990; Van Oostendorp & Kok, 1990), the effect of the position of the incorrect name in occurrences of semantic illusions (e.g., Bredart & Modolo, 1988; Van Jaarsveld et al., 1997), and the relation between syntactic structure complexity and the illusions (Gomes & Maia, 2024). Overall, most studies support the results found by Erickson and Mattson (1981). However, research on the impact of semantic relatedness in occurrences of semantic illusions also led to the postulate that participants seem to create a complete representation of the information presented by the question but are not able to match the information with the representation available in memory. Thus, the illusion question itself might have been encoded adequately, and all relevant information might have been retrieved, but the mismatch between the retrieved information and the information presented in the question remains unnoticed (e.g., Van Oostendorp & de Mul, 1990; see also partial-match hypothesis or imperfect memory match hypothesis postulated by Reder & Kusbit, 1991, for more details).

There are more recent hypotheses that consider a differential, although related, cognitive mechanism underlying the semantic illusion effect, namely the Node Structure Theory (NST) developed by Shafto and MacKay (2000). This model consists of a vast network of interconnected representational units called *nodes*, organized in semantic and phonological systems (see Figure 1).

Taking Figure 1 into account, the participant might notice the inconsistent name *Moses* (i.e., the phonological input) when the Moses-Illusion is presented, but additional information might already be activated due to shared nodes with the correct word. For instance, information such as *biblical characters* and *Old Testament* might be primed. Information about the event of building the Ark and about the number of animals of each kind is given by the question (i.e., the semantic input). According to this model, *Noah* receives more *priming* than *Moses* and becomes

https://doi.org/10.1017/langcog.2024.51 Published online by Cambridge University Press

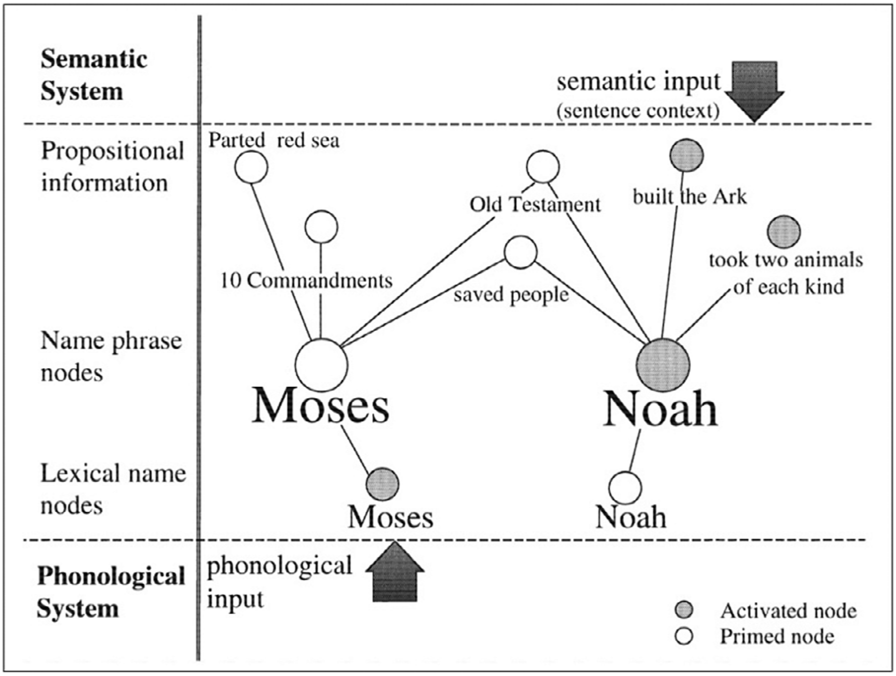

Figure 1. The Node Structure Theory (NST) explained for the Moses-Illusion.
Source: Shafto and MacKay (2000, p. 375).

activated under the *most-primed-wins-principle.* Consequently, participants do not notice the inconsistent name *Moses.*

Shafto and MacKay (2000) propose that not only semantic similarity (as also stated in previous research confirming the semantic similarity hypothesis, proposed by Erickson & Mattson, 1981) but also phonological similarity might cause participants to fall for semantic illusions. In the first experiment, participants heard four versions of experimental questions: a valid version, which contained the appropriate name, and three invalid versions containing names that were semantically related, phonologically related or unrelated to the valid name. A partial shadowing task was used, demanding special attention to critical words because participants read a written version of each question containing blank slots and had to shadow whatever auditory word occupied each blank slot. Findings indicated that semantic relatedness between the inconsistent name and the correct word might not be the only explanation for occurrences of semantic illusions, ruling out purely semantic theories (e.g., Erickson & Mattson, 1981; Van Oostendorp & Kok, 1990). Thus, the phonological similarity of words seems to play an important role in the semantic illusion effect as well. As an example of an illusion caused by phonological similarity, Shafto and MacKay (2000) presented the *Armstrong-Illusion.* After being exposed to the question, '*What was the famous line uttered by Louis Armstrong when he first set foot on the moon?*', participants usually understand this question as being valid, even though they know that Louis Armstrong was a jazz musician and that the first man landing on the moon was, in fact, Neil Armstrong. However, the exclusion of purely semantic

https://doi.org/10.1017/langcog.2024.51 Published online by Cambridge University Press

theories does not account for purely phonological theories of the Armstrong effect. More specifically, three summated factors, one phonological (i.e., the name *Armstrong*) and two semantic (i.e., information about being an *astronaut* and the *first man on the moon*), caused the unrelated name (i.e., *Neil Armstrong*) to accumulate more *priming* than the presented name (i.e., *Louis Armstrong*) and to become activated.

Some parallels can be found between the NST and the so-called 'Good Enough' approach (see Blott et al., 2021; Huang & Ferreira, 2021), according to which readers are not always aware of semantic ambiguities, leading to misinterpretations of the read sentences. This occurs because readers resort to shallow processing to be more efficient. Consider semantic garden-path sentences, like '*Sally worried that the ball was going to be too crowded for her liking*'. Since *ball* admits two semantic representations (the toy and the dancing event), two interpretations are possible and both are activated even when one of them may be more dominant than the other. The theory predicts that, in such a scenario, readers create a superficial linguistic representation of the meaning behind the presented sentence and consider it 'good enough' (Ferreira, 2003). The *good-enough approach* could also account for the occurrence of semantic illusions. Since both, the anomalous word and the correct word, share semantic representations, the participant might not be fully aware of the illusion. Thus, parallel to the *most-primed-wins-principle*, the sentence illustrated above offers information such as *too crowded* that might lead to more *priming* of the dancing event than of the toy.

The theories mentioned thus far provide a general understanding of semantic illusions. Overall, they agree with the idea that semantic illusions seem to occur due to some kind of semantic overlapping between the critical words. However, none of them can explain why the semantic illusion effect is modulated by factors such as the location of the erroneous word within the question or statement, formulation of instructions (e.g., Van Jaarsveld et al., 1997), the font of the written illusion (e.g., Song & Schwarz, 2008), and individual differences like reading fluency, long-term and working memory capacities (see Hannon, 2014 for a critical overview). It seems, therefore, that different cognitive mechanisms are involved in the failure to detect semantic anomalies, or as Hannon (2014) stated, multiple sources of misinformation lead people to fall into illusions.

1.2. Semantic illusions in second language processing

The picture is even more complex when participants have to deal with more than one language. The question of whether the effect also appears during second language (L2) processing has not yet been fully explored and has guided the present research. This question may seem trivial at first; however, its answer has important implications for models of bilingual processing that assume that age of acquisition (AoA), and the degree of language proficiency can influence the way lexical access occurs, and consequently, how sentences are processed in the L2 (the Multilink model, Dijkstra et al., 2019). Besides, studies report differences in types of processing that characterize the L1 and the L2 (a more intuitive processing for the L1 *vs.* a more controlled processing for the L2; Costa et al., 2014; Keysar et al., 2012) that may influence the semantic illusions effect. To the best of our knowledge, only a handful of studies have examined semantic illusions in bilinguals, and results are inconclusive

https://doi.org/10.1017/langcog.2024.51 Published online by Cambridge University Press

(i.e., Bautista Martín, 2022, unpublished master's thesis; Dhaene et al., 2022; Vaessen, 2017, unpublished master's thesis; see also Geipel et al., 2015).

Vaessen (2017) examined the effect of cross-linguistic differences (L1 *vs.* L2) and L2 proficiency on semantic illusions in late sequential Dutch-English bilinguals. The author expected that lower levels of English (L2) proficiency would have a negative effect on occurrences of semantic illusions (i.e., a higher proportion of semantic illusions) due to shallow sentence processing. In the first experiment, sentences were presented in an auditory manner. Four different conditions were tested: semantic illusions in L1 (condition 1), correct sentences in L1 (condition 2), semantic illusions in L2 (condition 3) and correct sentences in L2 (condition 4). Participants had to judge whether aurally presented questions were correct, incorrect or whether they did not know. Findings showed a significantly higher rate of semantic illusions in English (L2) than in Dutch (L1). However, no influence of L2 proficiency on semantic illusions was found in both assessed languages.

The second experiment consisted of a Rapid Serial Visual Presentation (RSVP) task accompanied by a scale for acceptability judgment. In this task, sentences were presented visually word by word instead of aurally, meaning that participants were unable to look back at the critical part of the sentence. No differences were found in the rate of illusion occurrences and in response times between L1 and L2. However, slight differences in response times were observed, indicating that bilinguals were faster to respond to semantic illusions in L1 than in their L2. Again, the level of L2 proficiency did not influence the overall results.

A recent eye-tracking study by Dhaene et al. (2022) found similar results to those of Vaessen (2017). The authors postulated that the cognitive load hypothesis could account for the mechanisms underlying semantic illusions. They examined the effect of semantic illusions in late sequential high proficient Dutch (L1)-English (L2) bilinguals. Anomalous and non-anomalous questions were presented on a screen and participants had to respond to the questions aloud. Results showed a higher proportion of illusions and slower reading times in the L2 than in the L1, sustaining the above-mentioned hypothesis. Accordingly, the cognitive load would cause a lower detection rate of illusions for the L2 than for the L1 (and thus, a high percentage of semantic illusions), probably because of a partial matching of the semantic information (as discussed by Reder & Kusbit, 1991). The opposite scenario would be expected if L2 processing was characterized by a more controlled processing, as Costa et al. (2014) hold.

Other studies have also made important contributions to this research domain. However, their research goals were not to intentionally study the semantic illusion phenomenon. For instance, Geipel et al. (2015) aimed to examine the foreign language effect (i.e., modulations in the cognitive processes responsible for judgment and decision-making when thinking in a non-native language *vs.* native language) in late but highly proficient sequential Italian (L1)-GER (L2) bilinguals. They conducted a series of experiments, in which the third one is of special interest for the present research. In this experiment, the Moses-Illusion was evaluated in both languages to test whether a foreign language promotes analytical or controlled reasoning (contrary to the more intuitive reasoning that characterizes L1 processing), as previous studies suggested (Costa et al., 2014; Keysar et al., 2012). Findings showed a similar rate of semantic illusions regardless of language, and thus the L2 did not improve performance on the Moses-Illusion task. Again, this conclusion is not consistent with the idea that an L2 promotes a switch from intuitive to more controlled processes (e.g., Costa

https://doi.org/10.1017/langcog.2024.51 Published online by Cambridge University Press

et al., 2014; Keysar et al., 2012), but rather suggests that intuitive processes remain present regardless of the tested language.

To summarize, results from the above-mentioned studies, although not conclusive, seem to suggest a similar processing regardless of the language (L1 or L2). If differences across languages appear, these are explained in terms of the cognitive load hypothesis. Thus, the aim of the present research is to present further evidence in this line of research by examining the effect of semantic illusions in EP (L1)-GER (L2) bilinguals who vary in the age of L2 acquisition and L2 proficiency. A group of EP native speakers (without knowledge of GER) and another of GER native speakers (without knowledge of EP) were also tested to assess the suitability of materials to provoke semantic illusions and, if so, to compare the proportion of semantic illusions across populations. Participants were asked to read sentence-formed illusions instead of illusion questions. This methodological choice was based on recent work done with native speakers of Brazilian Portuguese, showing evidence that sentences with semantic illusions, despite being anomalous, are perceived as well-formed sentences (Gomes & Maia, 2024).

Having taken into consideration tenets of the above-discussed theories on native speakers, and more specifically, those raised with bilinguals, such as the cognitive load hypothesis (Dhaene et al., 2022), we expected to find no differences in the effect of semantic illusions across languages in early bilinguals with balanced language proficiency (as Bautista Martín, 2022, has recently found). For late bilinguals, differences may, however, be observed between the languages in the form of more semantic illusions in L2 than in L1 (as well as slower reaction times), in line with postulates of the cognitive load hypothesis and consistent with what Dhaene et al. (2022) and Vaessen (2017) observed, especially for participants with lower degrees of L2 proficiency.

Language proficiency may, besides, blur the picture around age of language acquisition (AoA), since the L1 is not always the dominant language of bilingual speakers (Flores et al., 2022). This is an important issue to examine since there is evidence that L2 proficiency and age of acquisition contribute independently to bilinguals' comprehension and production behavior (García-Pentón et al., 2016; see also Bonfieni et al., 2019; Sá-Leite et al., 2023). Therefore, variations in proficiency and their impact on the semantic illusion effect were explored in terms of dominance. Dominance describes the difference in proficiency between two languages spoken by bilinguals and it is closely related to proficiency. We decided to include dominance in the analysis of the semantic illusions effect because recent research showed that dominance might affect bilingual language processing more than language proficiency, analyzed independently for L1 and L2 (Sá-Leite et al., 2023).

2. Experiment 1

In the first experiment, we aimed to assess the appropriateness of our materials and to examine the occurrence rate of semantic illusions in both languages, EP and GER. Therefore, this first experiment was conducted with GER-native speakers and EP-native speakers only, serving as a control study for the second experiment with EP-GER bilinguals.

Occurrences of semantic illusions were expected to be observed in both groups of participants, i.e., slower responses and more errors to sentences with anomalous

https://doi.org/10.1017/langcog.2024.51 Published online by Cambridge University Press

words, such as '*After the plane crash, the authorities* *buried* ***the survivors*** *in the mountains of the Alps*', than to correct sentences, '***the dead'***, because of shared semantic nodes between correct (*dead)* and anomalous words (*survivors)*, as observed in previous studies (e.g., Erickson & Mattson, 1981; Gomes & Maia, 2024; Shafto & MacKay, 2000).

3. Method

3.1. Participants

A total of 62 participants were tested – 27 were native speakers of GER (*Mage* = 31.3; *SD* = 12.9) of which 16 were females (one did not want to say and one chose 'other'), and 35 were native speakers of EP (*Mage* = 20.9; *SD* = 6) of which 30 were females (two did not want to say). All participants had normal or corrected-to-normal vision, normal hearing and were naïve about the main purpose of the study.

The experiment was carried out on the *web-based* Platform PCIbex (Zehr & Schwarz, 2018) to promote greater adherence to the study and to diminish the personal contact due to the pandemic situation of Covid-19. Participants were invited to participate through social networks (e.g., Facebook and Instagram) and personal contacts.

For the EP group, the same strategy was used, but EP students could also access an *accreditation platform*, which allowed them to improve their grades during the current semester through course credits after completing all the tasks of the study.

3.2. Materials

3.2.1. Stimuli

The task comprised 52 experimental sentences and 52 filler sentences in each language. Seventy percent of the sentences were taken from previous studies and translated into GER and EP by proficient EP and GER-speaking researchers (Barton & Sanford, 1993; Bredart & Modolo, 1988; Gomes & Maia, 2024; Koornneef & Reuland, 2016; Muller et al., 2020; Shafto & MacKay, 2000; Vaessen, 2017), whereas others were newly created following the structure of the former. Seventeen of the experimental sentences were of the type: '*After the plane crash, the authorities buried the survivors in the mountains of the Alps*' (Gomes & Maia, 2024), and 35 experimental sentences were based on general knowledge as in: '*After the great flood, Moses brought the animals in pairs to the Ark*'.

In this type of sentences, the post-verbal element could be an anomalous element, such as '*the survivors*' or a correct element '*the dead*'. Sentences were all grammatical, however, conceptual combination between verb and anomalous complement was not plausible. All sentences were manipulated with the goal that the first clause had a specific context, such as *after the plane crash* that should *prime* a specific verb-complement relationship, such as *bury-dead.* The control condition was composed of sentences in which the context *primed* the critical word. The semantic illusion condition was composed of control sentences, but with the verbal complement replaced by an anomalous complement, such as *survivors.* All sentences were divided into six segments and segments were controlled for length – one to eight words. Additionally, we tried to control for sentence structure, as best as possible. For EP, the sentences were composed of the following segments: *sentence context – subject – verb – critical segment – rest of the sentence.* GER sentences followed EP sentence structure,

https://doi.org/10.1017/langcog.2024.51 Published online by Cambridge University Press

but the subject and the verb had to be switched due to the German verb second property. Thus, GER sentences had the following structure: *sentence context – verb – subject – critical segment – rest of the sentence.*

Additionally, corresponding target words in the control and illusion conditions were controlled for their frequency (logarithmic frequency, *ps* > .69) and length (all *ps* > .74) in each language. However, it is important to note that item length and sentence length showed differences across languages (*ps* < .01), due to the fact that GER words and phrases were generally longer than those in EP. Frequency values were obtained from the SUBTLEX-DE database (Brysbaert et al., 2011) and the SUBTLEX-PT database (Soares et al., 2015). Words were also controlled for their phonological degree of similarity (NLD). The NLD between the anomalous and correct EP words was equal to 0.69 and the NLD between GER words was 0.71 (*p* = .109). Values for phonological NLD were taken from the PHOR-in-One database (Costa et al., 2022).

A total of 104 sentences (52 experimental + 52 distractor) were created in both languages, EP and GER, in a 2 × 2 Latin square experimental design (two conditions per two languages), as in Table 1.

Across conditions, materials following the critical segment (*the survivors* vs. *the dead*) were kept the same: an adverbial phrase, such as *in the Alps.* This manipulation affected the material after the critical word and does not directly affect responses to this task on the processing of the critical segment. The use of this element is also intended to compensate for *spill-over effects* widely described in literature in self-paced reading tasks (see Jegerski, 2014).

We employed a self-paced Rapid Serial Visual Presentation (RSVP) task with a moving window, wherein the reading segment disappears as soon as the next segment is revealed. Consequently, there is no possibility of returning to the previous segment. Thus, participants were forced to read one segment at a time and could not read the word before or after the critical word. However, participants could continue to process a segment as they read subsequent segments. This phenomenon is described in the literature as *spillover effect*: the difficulty induced by a particular segment could slow down the reading/processing time of one or more subsequent segments (Koornneef & van Berkum, 2006; Mitchell, 1984; Smith & Levy, 2013).

From these materials, four lists were created: two lists of stimuli in each language (one list with control sentences in EP and one in GER) and two lists of illusion

Table 1. Latin square: Condition vs. item

Control EP	Depois do acidente aéreo *After the accident*	as autoridades enterraram ***os mortos*** *the authorities buried the dead*	nos Alpes *in the Alps*
Control GER	Nach dem Unfall *After the accident*	begruben die Rettungskräfte ***die Toten*** *buried the rescue workers the dead*	in den Alpen Gebirgen *in the Alps mountains*
Illusion EP	Depois do acidente aéreo *After the accident*	as autoridades enterraram ***os sobreviventes*** *the authorities buried the survivors*	nos Alpes *in the Alps*
Illusion GER	Nach dem Unfall *After the accident*	begruben die Rettungskräfte ***die Überlebenden*** *buried the rescue workers the survivors*	in den Alpen Gebirgen *in the Alps mountains*

https://doi.org/10.1017/langcog.2024.51 Published online by Cambridge University Press

sentences (one in EP and one in GER). Each participant read only one sentence of each type, half of the experimental phrases contained the illusion term, while the other half contained the correct term. Thus, a total of 104 sentences (52 experimental + 52 fillers) comprised each list in both languages. In addition, each list containing experimental sentences (control and illusion) and fillers was pseudo-randomized, respecting the criterion of including at least one distractor item among experimental items. Four lists were randomly presented to participants. This design allowed for testing the same materials in both control and illusion paradigms (see Table 1) in a within-subjects paradigm.

3.2.2. General knowledge test

Additionally, eight lists were created for the general knowledge test on *Google Forms* (Google, 2022) for each condition and each language. The *Google Form* contained semantic illusions that had been previously shown to participants and required general knowledge. The test was designed as a multiple-choice test in which participants had to indicate whether they knew the answer or not. For example, when presented with a general knowledge sentence such as '*After the great flood, Moses brought the animals in pairs onto the Ark*', a question focusing on the target followed: '*Did Moses bring the animals onto the Ark?*'. Participants were given three possible choices: '*yes*', '*no*' or '*don't know*'. General knowledge was tested as a prerequisite for a semantic illusion to occur; participants must have some world knowledge about the information presented. If the complete semantic concept is unfamiliar, participants would not fall for the illusion (Van Oostendorp & Kok, 1990).

3.2.3. Language experience and proficiency assessment

A sociodemographic questionnaire and the standardized DIALANG Vocabulary Size Placement Test (see Alderson, 2005) were used to assess participants' language experience and proficiency. The questionnaire included questions about participants' age, sex, education level, and knowledge of any second language. The standardized DIALANG Vocabulary Size Placement Test (see Alderson, 2005) is a lexical decision task in which participants must decide if a certain sequence of letters exists as a word in the given language. Monolingual controls completed the task in their native language – EP or GER. Lexical competence has been argued to be a valid predictor of language proficiency (Laufer & Nation, 1999; Treffers-Daller & Korybski, 2016), because speakers' lexical inventory grows as proficiency increases. The DIALANG Vocabulary Size Placement Test assesses lexical knowledge through a list of 75 items, including 50 real words and 25 non-words. Following Alderson's (2005) proposal, participant's scores are computed based on the total number of words correctly identified as either real words or non-words, with 1 point awarded per correct assessment (see also Flores et al., 2022).

3.3. Procedure

Before starting the experiment, all participants were asked to sign the online informed consent form, which aimed to explain the main objectives of this study and the confidentiality of participants' data. Upon accepting to participate in the study, participants were directed to perform the self-paced RSVP. One of the advantages of the self-monitoring reading paradigm is its compatibility with web-based platforms. This study was hosted on the web-based platform PCIbex – PennController for Internet-Based Experiments (Zehr & Schwarz, 2018). Participants were tested in a single online session lasting approximately 15 minutes (inclusive of the informed consent and instructions).

Prior to the experimental session, participants needed to complete a training session in their native language with 10 sentences each. As in a self-paced reading experiment, participants determined the speed at which sentence segments were presented on the screen by pressing the *space* button, allowing for the measurement and evaluation of reading times (RTs) for each segment. Each segment was presented in a moving-window paradigm (SPR-moving-window) at the center of the screen.

Each experiment consisted of the following events: A sentence trace is revealed, segment by segment, in the center of the screen. After the participant pressed the *space* button, the next segment is revealed until they read the entire sentence. The time a segment is visible is used as a dependent measure and reflects reading time (RT). After reading the entire sentence, a question appears on the screen: *Does it make sense?* Participants must then decide whether the read sentence makes sense in EP or GER. They were instructed to answer, '*Yes*', if the sentence was semantically coherent and grammatically well-formed, and 'No' otherwise. Participants answered the questions by pressing one of two buttons, which were counterbalanced (left and right) between participants. Each participant was randomly assigned to a counterbalanced list of stimuli and was instructed to take the test on a computer, while reading as normally as possible, trying to understand the sentences, and answering the questions as accurately as possible. An 85% accuracy threshold on globally plausible items in the behavioral task was set to ensure that participants were paying attention.

Finally, participants were asked to respond to the sociodemographic questionnaire, the general knowledge test and the standardized DIALANG Vocabulary Size Placement Test. Links to the general knowledge test (corresponding to the list they were assigned to for the self-paced RSVP) and the standardized DIALANG Vocabulary Size Placement Test were sent to each of the participants after they completed the sociolinguistic questionnaire.

4. Results

4.1. Results for 'yes' responses

Error rates (%E) and reading times (RT) in ms obtained for the first segment of the sentences, in which the anomalous word or correct word appeared (RT1), and for the segment following the first one (RT2) were analyzed using linear mixed-effect models (e.g., Baayen, 2008; Baayen et al., 2008). The analysis of the second segment was included in order to examine *spillover effects.* It is important to note that, in the RT analyses, we included correct responses to control items ('*yes*' responses) and errors to semantic illusion items ('*yes*' responses). To this end, the lme4 package of R was used (Bates et al., 2015). A series of fixed structure models were created to examine the hypotheses of the study, with the %E or RTs as dependent variables. As fixed effects, models included Condition (control or semantic illusion) and Length (number of letters of the item) for native speaker analyses. Length was centered and transformed into *Z*-scores. To control for variability due to participants and

https://doi.org/10.1017/langcog.2024.51 Published online by Cambridge University Press

items (including the type of illusion) involved in the experiment, we introduced participants and items as random effects in the models. As items could appear in two different conditions (control or semantic illusion), we also included a slope of Condition in the random effect of items. We were unable to increase the complexity of the random structure of the models as this would not allow them to converge. The formula of the models was the following: (%E or RT) ~ condition + length + (1 | participant) + (condition | item).

The significance of effects and interactions was determined using log-likelihood ratio tests (R anova function). The contribution of each effect and interactions were assessed by comparing a model that included them with another model in which they were not included. Also, results of *t*-test analyses for coefficient estimates of fixed effects are reported. To this end, Satterthwaite's approximations to the degrees of freedom of the denominator were used (*p*-values were estimated by the lmerTest package; Kuznetsova et al., 2017).

4.1.1. EP group

Prior to the analyses, some data filtering was conducted to remove observations with RTs below 300 ms or above 3000 ms, as well as observations corresponding to general knowledge items for which the participant did not know the answer. This resulted in the removal of 8.22% of total observations. No participant or item was excluded from the analyses. See Table 2 for mean RT and %E of the experiment.

Results revealed no significant effect ($p > .05$) of item (between straight relation verb-complement and contextual semantic illusion) and no significant effect of Condition on %E (*estimate* = 0.43, *SE* = 0.28, 95% CI [−0.14, 1.03], $z = 1.52$, $p = .129$, $\chi^2(1) = 2.26$, $p = .133$). However, the number of semantic illusions was slightly higher than the errors for control sentences (i.e., difference of 4.8%). In terms of RT1s, there was also no significant effect of Condition on the critical segment (*estimate* = 38.90, *SE* = 31.83, 95% CI [−26.49, 100.86], $t = 1.22$, $p = .227$, $\chi^2(1) = 1.50$, $p = .220$). RT1s for control sentences were found to be only slightly lower than for semantic illusions (i.e., difference of 28 ms). In contrast, when considering RT2s (i.e., RTs for a segment following the critical one), results showed a significant effect of Condition (*estimate* = 94.09, *SE* = 28.78, 95% CI [35.55, 150.40], $t = 3.27$, $p = .002$, $\chi^2(1) = 10.18$, $p = .001$), indicating significant slower RTs in semantic illusions compared to control sentences (i.e., difference of 59 ms).

4.1.2. GER group

The same data filtering procedure as the one used for EP native speakers was applied, resulting in the removal of 7.26% of total observations. No participant or item was excluded from the analyses. See Table 3 for mean RT and %E of the experiment.

Table 2. Mean RT and %E of each experimental condition for the EP group

Condition	%E	RT1	RT2
Control	24.1 (1.50)	682 (15.6)	670 (12.9)
Illusion	28.9 (1.59)	710 (23.3)	729 (24.8)

Note: RT1 shows the mean RT for the critical segment, whereas RT2 shows the mean RT for the segment following the critical one. Standard errors are included in parentheses.

https://doi.org/10.1017/langcog.2024.51 Published online by Cambridge University Press

Table 3. Mean RT and %E of each experimental condition for the GER group

Condition	%E	RT1	RT2
Control	17.6 (1.50)	824 (16.6)	979 (21.5)
Illusion	33.0 (1.84)	864 (29.2)	990 (28.9)

Note: RT1 shows the mean RT for the critical segment, whereas RT2 shows the mean RT for the segment following the critical one. Standard errors are included in parentheses.

Table 4. Mean RT and %E of each experimental condition for the EP group

Condition	%E	RT1	RT2
Control	24.1 (1.50)	682 (15.6)	670 (12.9)
Illusion	28.9 (1.59)	737 (18.5)	779 (18.2)

Note: RT1 shows the mean RT for the critical segment, whereas RT2 shows the mean RT for the segment following the critical one. Standard errors are included in parentheses.

Table 5. Mean RT and %E of each experimental condition for the GER group

Condition	%E	RT1	RT2
Control	17.6 (1.50)	824 (16.6)	979 (21.5)
Illusion	33.0 (1.84)	920 (24.6)	1029 (23.8)

Note: RT1 shows the mean RT for the critical segment, whereas RT2 shows the mean RT for the segment following the critical one. Standard errors are included in parentheses.

In the GER group, results indicated a significant effect of Condition on %E. Participants committed more errors in the illusion condition than in the control condition (*estimate* = 1.15, *SE* = 0.36, 95% CI [0.41, 1.89], $z = 3.20$, $p = .001$, $\chi^2(1) = 9.60$, $p = .002$), showing a difference of 15.4%. Regarding RT1s, the effect of Condition on the critical segment was not significant (*estimate* = 21.95, *SE* = 36.74, $t = 0.60$, 95% CI [−50.21, 98.42], $p = .553$, $\chi^2(1) = 0.37$, $p = .545$), suggesting that control sentences and semantic illusions yielded close RT1s. Still, participants responded slightly slower to the illusion sentences than to the control sentences (i.e., difference of 40 ms). Additionally, the effect of Condition on the segment following the critical segment was also non-significant (*estimate* = −8.23, *SE* = 28.46, 95% CI [−71.88, 42.10], $t = 0.29$, $p = .774$, $\chi^2(1) = 0.08$, $p = .773$), further suggesting that there were no differences in RT2s between control sentences and semantic illusions in this segment as well. Yet, again, participants responded slightly slower to the illusion sentences than to the control sentences (i.e., difference of 11 ms).

4.2. Results for correct responses

We also ran the analyses on RTs for correct responses to control items ('*yes*' response) and correct responses to semantic illusion items ('*no*' response). See Tables 4 and 5.

4.2.1. EP group

In terms of RT1s, there was a significant effect of Condition on the critical segment (*estimate* = 57.82, *SE* = 22.99, 95% CI [12.01, 104.29], $t = 2.52$, $p = .016$, $\chi^2(1) = 5.95$,

https://doi.org/10.1017/langcog.2024.51 Published online by Cambridge University Press

p = .015). RT1s for semantic illusions were slower in comparison to control sentences (i.e., difference of 55 ms). When considering RT2s (i.e., RTs for a segment following the critical one), results also showed a significant effect of Condition (*estimate* = 105.03, *SE* = 25.06, 95% CI [50.29, 144.15], t = 4.19, p < .001, $\chi^2(1)$ = 15.01, p < .001), indicating slower RT2s in semantic illusions compared to control sentences (i.e., difference of 109 ms).

4.2.2. GER group

Regarding RT1s, the effect of Condition on the critical segment was significant (*estimate* = 102.52, *SE* = 29.76, 95% CI [45.37, 164.26], t = 3.45, p = .001, $\chi^2(1)$ = 10.61, p = .001), showing slower RT1s for semantic illusions than for control sentences (i.e., difference of 96 ms). In contrast, the effect of Condition on the segment following the critical segment was non-significant (*estimate* = 53.42, *SE* = 32.96, 95% CI [−11.45, 120.39], t = 1.62, p = .112, $\chi^2(1)$ = 2.59, p = .107). However, participants responded slower to the semantic illusions when compared to the control sentences (i.e., difference of 50 ms).

5. Discussion

The primary goal of the first experiment was to examine the suitability of the materials used in eliciting the effect of semantic illusions on %E and RTs in EP and GER native speakers, and if so, to determine if semantic illusions occur in both languages. For the group of EP speakers, it was observed that participants did not make significantly more errors in the illusion condition than in the control condition, contrary to initial expectations. However, it is important to note that %E was numerically higher for the illusion condition compared to the control condition (4.8% of semantic illusions, see Tables 2 and 4). Yet, it is essential to emphasize once again that a condition effect was observed for RT2. The effect was more evident for GER speakers – participants committed significantly more errors in the illusion condition when compared to the control condition (13.4% of semantic illusions). In contrast to EP, however, no significant differences in RTs (neither RT1 nor RT2) between both conditions were shown, although RTs were slightly slower for the illusion condition. As for correct responses to control items ('yes' response) and correct responses to semantic illusion items ('no' response), it is important to note that RTs showed a condition effect for EP and GER, suggesting that participants needed more time to respond accurately to illusion sentences than to control sentences.

As a possible explanation for why the effect was nearly three times larger in GER speakers than in EP speakers, we propose that the robust semantic illusion effect is influenced by participants' reading speed, as GER participants were slower than EP participants (see Tables 2 and 3, and Tables 4 and 5). Consequently, more room may exist for the effect of semantic illusions to manifest. However, an analysis of the correlation between the mean of RT and %E showed a non-significant relationship (r = −.19; p = .136). Hence, there seems to be another explanation behind these results that cannot be solely attributed to participants' reading speed. This issue will be further addressed in the General Discussion.

6. Experiment 2

The second experiment aimed at examining the occurrence of semantic illusions in each of the two languages (EP and GER) spoken by the early and late EP-GER bilinguals. In particular, %E and RTs to each semantic illusion phrase were assessed.

As mentioned earlier, no differences between the two languages were expected for early bilinguals (see also Bautista Martín, 2022), while a higher occurrence rate of semantic illusions in the L2 than in the L1 was predicted for late bilinguals (as found by Dhaene et al., 2022; Vaessen, 2017). Furthermore, it was expected that no impact of dominance would be observed, given previous work showing that semantic illusions occur independently of the level of proficiency (see Vaessen, 2017).

7. Method

7.1. Participants

Data of 40 bilingual speakers of EP and GER with a mean age of 36.9 years (SD = 11.2), 33 females, were collected. As inclusion criterion, participants had to have normal or corrected-to-normal vision, normal hearing and had to be naïve about the objective of the study. In addition, they had to be able to maintain a fluent conversation in EP and GER.

All 40 bilinguals were L1 speakers of EP. Twenty-one acquired GER in childhood in a naturalistic context, before the age of 10, being early sequential bilinguals. Nineteen were late learners, having acquired GER as teenagers in a classroom context. More information on the speakers' sociolinguistic background was collected in a sociodemographic questionnaire and the standardized DIALANG Vocabulary Size Placement Test, and is available in Table 6 (see Section 7.2).

7.2. Materials

7.2.1. Stimuli

Stimuli remained the same as those used in the first experiment. Since the second experiment involved testing bilingual participants, it was necessary to create eight experimental lists: the four lists used in the first experiment, but now in both EP and GER.

7.2.2. General knowledge test

The same *general knowledge tests* were used. Bilingual participants completed the tests in their L1.

7.2.3. Language experience and proficiency assessment

In addition to the same questions asked in Experiment 1, information on the age of onset of bilingualism and place of residence was collected. Participants were also asked to self-rate their proficiency in reading, writing and speaking in each of the languages on a scale from 1 to 7 (*Likert Scale*) and to indicate the percentage of time they spend per day using each of the languages in a week. This allowed us to examine the age of L2 acquisition, participants' subjective proficiency, and degree of use for both languages. Finally, they were asked whether they have ever lived abroad and if

Table 6. Summary of the bilinguals' background variables and DIALANG results

Variable		Early learners	Late learners
Age	Mean (*SD*)	38.1 (12.0)	34 (9.9)
	Min–max	16–55	20–50
AoA German	Mean (*SD*)	2.4 (2.9)	19.8 (5.6)
	Min–max	0–10	13–31
Current contact with German	Mean (*SD*)	30.2 (19.6)	30.4 (21.9)
	Min–max	0–75	0–75
Current contact with Portuguese	Mean (*SD*)	58.3 (22.4)	50.7 (23.3)
	Min–max	20–90	20–85
Self–evaluation German	Mean (*SD*)	12.3 (1.6)	9.9 (2.8)
	Min–max	9–14	4–13
Self–evaluation Portuguese	Mean (*SD*)	12.7 (1.1)	13.5 (0.9)
	Min–max	11–14	12–14
DIALANG German	Mean (*SD*)	67 (6)	55.7 (6.9)
	Min–max	52–75	45–70
DIALANG Portuguese	Mean (*SD*)	64.9 (3.7)	65.7 (3.1)
	Min–max	60–71	62–71
DIALANG Dif	Mean (*SD*)	2.1 (7.4)	−10 (5.7)
	Min–max	−17–12	−19–0

Note: The total score of self-evaluation is the sum of the speaking and the writing scores (maximal score: 14). The difference between the score obtained in the GER and in the EP test (negative values indicate dominance in EP; positive values indicate dominance in GER; 0 equal dominance for EP and GER) is abbreviated as DIALANG Dif.

so, for how long, so that some insight about their linguistic environment throughout their lives could be obtained.

Furthermore, the same standardized DIALANG Vocabulary Size Placement Test (see Alderson, 2005) was administered twice – once in GER and a second time in EP, or vice versa. Table 6 summarizes the main background variables.

As can be seen in Table 6, early and late bilinguals presented similar rates of current contact with GER (early: 30.2 [19.6]; late: 30.4 [21.9]). A *t*-test showed no group differences regarding current contact with GER ($p = .98$). As expected, current contact with EP was higher compared to GER (early: 58.3 [22.4]; late: 50.7 [23.3]), but again no statistically significant group differences were observed ($p = .37$). This indicates that, at time of testing, both groups had similar profiles regarding the amount of contact with the tested languages. As for the difference in self-evaluation in GER, early bilinguals assessed their proficiency higher than late learners (early: 12.3 [1.6]; 9.9 [2.8]), reaching statistical significance ($p = .008$). The opposite was observed for self-evaluation in EP, with late bilinguals assessing their proficiency in EP higher than early bilinguals (early: 12.7[11]; late: 13.5[0.9]). This difference was also significant ($p = .038$). When considering the DIALANG values, the difference between early and late learners was significant for GER ($p < .001$), but not for EP ($p = .58$), as expected.

7.3. Procedure

The procedure for all tasks was similar to the one used in Experiment 1. The only difference was that bilinguals were solicited to conclude the self-paced RSVP task in both languages. Therefore, after signing the informed consent form, each participant did half of the self-paced RSVP in EP and half of the experiment in GER in a single session, in two blocks separated by a short break and a short message to make participants aware of the language changing. The order of the languages was counterbalanced – half of the participants started the experiment in EP and the other half started it in GER. Half of the sentences within each block contained the correct verbal complement (control), and the other half the incorrect one (illusion), following the same procedure as for native speakers. In addition, the experimental sentences (control and illusion) and fillers in each of the eight lists were pseudo-randomized. Additionally, the eight lists with fillers were randomly presented. Finally, participants were requested to respond to the sociodemographic questionnaire, the General Knowledge Test and the DIALANG Vocabulary Size Placement Test.

8. Results

The same data filtering procedure was applied as for the EP native speakers and GER native speakers, resulting in the exclusion of 5.05% of total observations from the analyses. No participant or item was excluded. For bilingual analyses, fixed effects were the triple interaction and second-order interactions between Condition (control or semantic illusion), Target Language (GER or EP) and Age of Acquisition (AoA; early or late). Language Dominance[1] and Length (number of letters of the item) were introduced as covariables (~ condition × target_lang × AoA + lang_dominance + length (1 | subject) + (condition | item)). Language Dominance and Length were centered and transformed into Z-scores. In addition, dichotomous variables were coded using sum contrast coding (−0.5 for the first level and +0.5 for the second level of each factor); Condition: control (−0.5), illusion (+0.5); Target Language: GER (−0.5), EP (+0.5); AoA: early (−0.5), late (+0.5). As in Experiment 1, we introduced participants and items as random effects in the models, and we also included a slope of Condition in the random effect of items. In the RT analyses, we included correct responses to control items ('yes' response) and error responses to semantic illusion items ('yes' response). See Table 7 for mean RT and %E of the experiment.

Analyses showed a significant interaction on %E between Target Language and Condition, $\chi^2(1) = 7.54$, $p = .006$. As can be seen in Figure 2, although the %E in the illusion condition was significantly higher than in the control condition in GER (*estimate* = 0.61, *SE* = 0.23, 95% CI [0.16, 1.06], $z = 2.67$, $p = .008$) and EP sentences (*estimate* = 1.27, *SE* = 0.23, 95% CI [0.83, 1.72], $z = 5.61$, $p < .001$), the effect was larger in EP sentences than in GER sentences. No other significant interactions were found in the analyses of %E. Note that RTs to GER stimuli were again slower than to their EP counterparts, although this finding was mediated by the interaction between target language and AoA (see Figures 3 and 4).

Data analysis on RTs to the critical segment (see Figure 3) showed an interaction between AoA and Target Language, $\chi^2(1) = 9.00$, $p = .003$, due to faster RT1s in late bilinguals compared to early bilinguals in EP (*estimate* = −270.2, *SE* = 111.0, 95% CI [−495, −45.1], $t = 2.44$, $p = .020$), but not in GER (*estimate* = −80.8, *SE* = 112.0, 95% CI

[1]The Language Dominance measure corresponds to the differences in DIALANG specified in Table 6 (DIALANG Dif.). We decided to consider this variable instead of scores per language (DIALANG GER – DIALANG EP) because it is easier to estimate the model if only one variable is introduced instead of two, especially when the number of observations is small, as was the case here. We also opted for including this objective variable into the model as an index of proficiency because objective and subjective measures of proficiency correlated positively here ($r = .64$; $p < .001$, for the GER scores) and in previous studies (see de Bruin et al., 2017; Marian et al., 2007).

https://doi.org/10.1017/langcog.2024.51 Published online by Cambridge University Press

Table 7. Mean RT and %E of each experimental condition for the bilingual group

Target Lang.	AoA	Condition	%E	RT1	RT2
German	Early	Control	28.9 (3.39)	1187 (44.6)	1306 (50.7)
		Illusion	32.0 (3.57)	1249 (81.4)	1467 (81.1)
	Late	Control	26.2 (3.10)	1158 (50.2)	1267 (47.9)
		Illusion	43.1 (3.43)	1202 (63.5)	1468 (69.2)
Portuguese	Early	Control	21.6 (3.16)	1043 (40.7)	1014 (37.8)
		Illusion	42.5 (3.70)	1178 (71.0)	1145 (56.1)
	Late	Control	26.3 (2.99)	856 (40.1)	821 (37.4)
		Illusion	53.5 (3.43)	951 (52.7)	826 (41.8)

Note: RT1 shows the mean RT for the critical segment, whereas RT2 shows the mean RT for the segment following the critical one. Standard errors are included in parentheses.

[−308, 146], $t = 0.72$, $p = .474$). This interaction was also observed in RTs of the segment following the critical segment (see Figure 4), $\chi^2(1) = 12.34$, $p < .001$, again with faster RT2s in late bilinguals compared to early bilinguals in EP (*estimate* = −277.9, *SE* = 105.0, 95% CI [−491, −64.6], $t = 2.65$, $p = .012$), but not in GER (*estimate* = −66.4, *SE* = 116.0, 95% CI [−282, 149.2], $t = 0.62$, $p = .536$). In addition, there was a significant Condition effect in RT2s (*estimate* = 95.65, *SE* = 37.98, 95% CI [21.11, 170.19], $t = 2.52$, $p = .015$), showing that RT2s for semantic illusions were slower than for control sentences.

As in Experiment 1, we conducted analyses for correct responses to control items ('yes' response) and correct responses to semantic illusion items ('no' response). See Table 8.

Data analysis on RTs to the critical segment showed an effect of Condition (*estimate* = 124.07, *SE* = 36.35, 95% CI [52.74, 195.39], $t = 3.41$, $p = .001$), suggesting

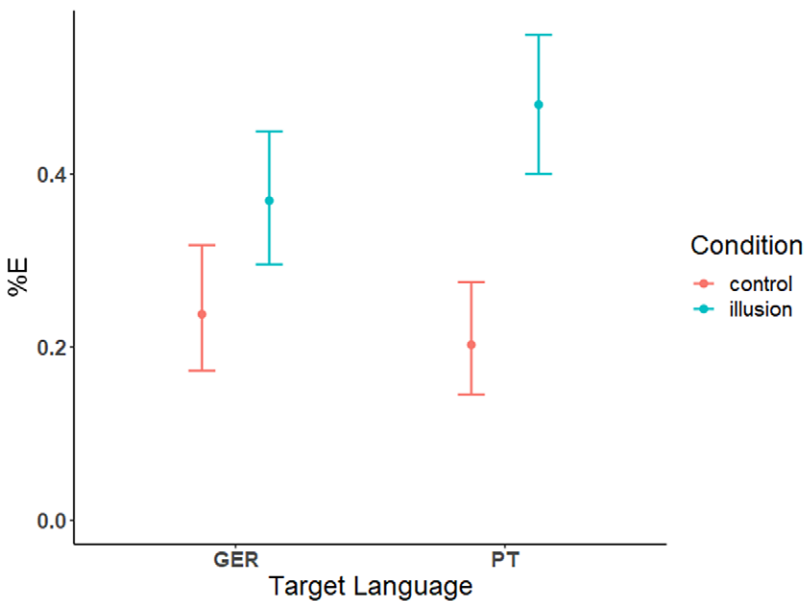

Figure 2. Interaction between target language (GER vs. Portuguese, PT) and condition (control vs. illusion) on %E.

https://doi.org/10.1017/langcog.2024.51 Published online by Cambridge University Press

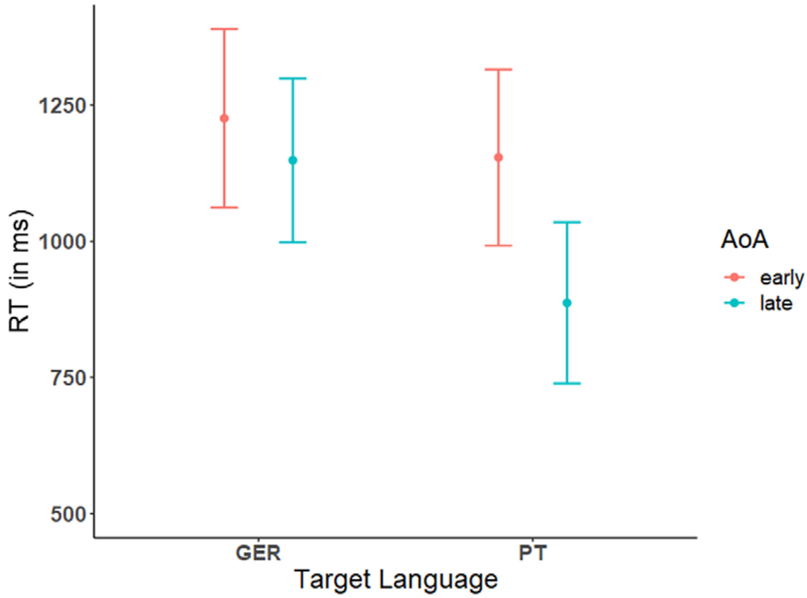

Figure 3. Interaction between target language (GER vs. Portuguese, PT) and AoA group (early vs. late) on RTs of the critical segment.

Table 8. Mean RT and %E of each experimental condition for the bilingual group

Target language	AoA	Condition	%E	RT1	RT2
German	Early	Control	28.9 (3.39)	1187 (44.6)	1306 (50.7)
		Illusion	32.0 (3.57)	1385 (61.4)	1290 (51.8)
	Late	Control	26.2 (3.10)	1158 (50.2)	1267 (47.9)
		Illusion	43.1 (3.43)	1273 (52.1)	1167 (56.5)
Portuguese	Early	Control	21.6 (3.16)	1043 (40.7)	1014 (37.8)
		Illusion	42.5 (3.70)	1081 (51.2)	1001 (51.2)
	Late	Control	26.3 (2.99)	856 (40.1)	821 (37.4)
		Illusion	53.5 (3.43)	913 (58.5)	935 (52.8)

Note: RT1 shows the mean RT for the critical segment, whereas RT2 shows the mean RT for the segment following the critical one. Standard errors are included in parentheses.

that RT1s for semantic illusions were slower than for control sentences. Analyses also showed a significant triple interaction between Group, Target Language and Condition on RT2s, $\chi^2(1) = 3.88$, $p = .049$. As can be seen in Figure 5, this interaction reflects an effect of semantic illusions which was restricted to EP sentences in the group of late bilinguals (see Figure 5), late bilinguals showed slower RT2s in the illusion condition than in the control condition (estimate = 144.9, $SE = 60.2$, 95% CI [−10.5, 300.3], t = 2.41, $p = .017$).[2]

[2]Following the suggestion of an anonymous reviewer, we decided to run post hoc power simulations for the primary analyses using the SIMR package in R (Green & MacLeod, 2016), in order to assess the influence of the rather small sample size on the results. The statistical power was not as high as desired for both the RT and %E models (between 24% and 87% for RT models and between 20% and 78% for %E models). Although we recognize that this is a limitation of the study and further research is needed, it is important to bear in mind the retrospective nature of this type of analysis.

https://doi.org/10.1017/langcog.2024.51 Published online by Cambridge University Press

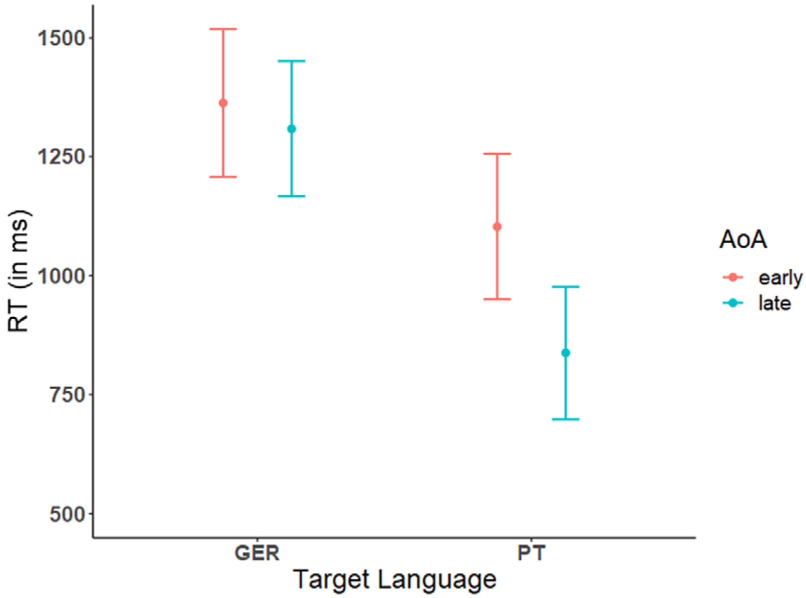

Figure 4. Interaction between target language (GER vs. Portuguese, PT) and AoA group (early vs. late) on RTs of the segment following the critical segment.

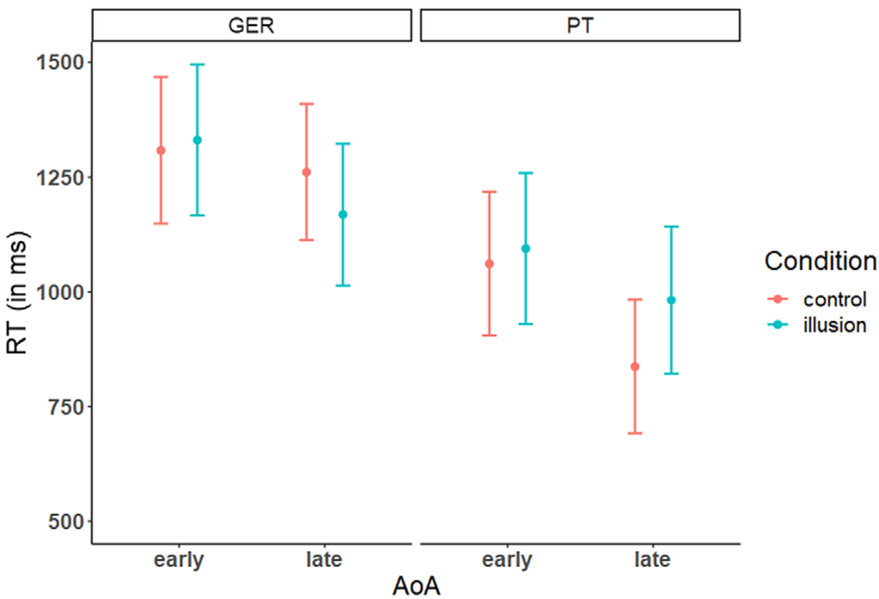

Figure 5. Interaction between condition, target language (GER vs. Portuguese, PT), and AoA group (early vs. late) on RTs of the segment following the critical segment.

https://doi.org/10.1017/langcog.2024.51 Published online by Cambridge University Press

9. Discussion

The aim of Experiment 2 was to examine the effect of semantic illusions in EP-GER bilinguals who vary in age of L2 acquisition and L2 proficiency. It was hypothesized that no differences would be observed for early learners of GER, but for late learners more semantic illusions were expected to occur in the L2 than in the L1 (as observed by Dhaene et al., 2022; Vaessen, 2017). Results were clear-cut: we observed an overall effect of SI in the %E in both languages (although larger in EP) regardless of group (early and late). When looking at RTs from 'yes' responses to correct and incorrect conditions (especially RT2s), results also showed that participants needed more time to respond to semantic illusions than to control sentences. Late bilinguals exhibited faster RTs in EP compared to early bilinguals, who acquired EP as L1 in a GER-dominant setting (as heritage language). This is explained by the fact that early bilinguals are less proficient in EP than late bilinguals, who were raised monolingual in an EP-dominant context. This difference in proficiency was evident from the self-evaluation and the DIALANG test results (see Section 7.2). Crucially, when RTs from correct responses (i.e., 'yes' responses to correct sentences and 'no' responses to incorrect ones) were considered (especially RT2s), the effect of SI was restricted to late bilinguals in their L1 (EP). No effects were observed in GER.

10. Overall discussion

The present study had the main goal to explore the effect of semantic illusions in early and late EP-GER bilinguals who vary in their age of onset of acquisition of GER and in their degree of language proficiency. Two groups of native speakers – GER native speakers and EP native speakers – worked as controls and were also tested in an attempt to assess the efficacy of materials to produce semantic illusions. Results showed a significant effect of semantic illusions on %E which was restricted to native speakers of GER. The pattern of results was similar for native speakers of EP but it was only significant for RT2. It is important to note that the percentage of semantic illusions observed in these two groups was similar to the results observed in Bautista Martín's (2022) study using the same procedure, namely around 15%. Likewise, differences were observed for the native groups between the semantic illusion condition and the correct condition when RTs were analyzed for correct responses to control items ('yes' response) and error responses to semantic illusion items ('yes' response), suggesting that participants took longer to process sentences with anomalous words than correct sentences.

Overall, results observed for native speakers suggest that semantic illusions influenced the %E, at least in GER, due to similarities between the anomalous and the correct word, supporting the semantic/phonological similarity hypothesis proposed by Shafto and MacKay (2000). Although the effect was not significant in EP, participants tended to commit more errors when confronted with sentences from the semantic illusion condition than when confronted with sentences from the control condition. Given that in the control group the percentage of illusions showed to be relatively small (first experiment with native speakers of EP), we suggest that the effect should also be explored with questions (as also proposed by Bautista Martín, 2022). The percentage of semantic illusions observed in previous studies using questions was substantially higher than the one found in the present study (ranging from 43% in the study by Erickson & Mattson, 1981, to 60% in the study by Dhaene et al., 2022).

https://doi.org/10.1017/langcog.2024.51 Published online by Cambridge University Press

Still, the reason why the effect was higher in the GER control group than in the EP control group is unclear. We tentatively hypothesize that participants' reading speed when reading in GER could be explaining the results because GER and EP have distinct language characteristics. However, correlation analyses between RTs and %E debunked this hypothesis, as above mentioned. Besides, results from the bilinguals show clear illusion effects in both languages, EP and GER.

In fact, a major finding of the present study is that early as well as late bilinguals show illusion effects in terms of %E in both their languages, independently of AoA, even though the effect is larger in EP than in GER. Furthermore, language dominance also did not influence the occurrence of illusions. When considering the analyses for correct responses to control sentences and incorrect responses to illusion sentences, early and late bilinguals only differ in their RT2s in EP, but not in GER. Additionally, when looking at correct responses to control sentences and correct responses to illusion sentences, a clear semantic illusion effect is shown for RT1s – participants need more time to respond to illusion sentences. This result in RT1s was independent of language and AoA. Yet, for RT2s, a triple interaction suggests that RT2s are slower for the illusion condition in EP, but only for late learners.

Thus, semantic illusions occur equally in L1 and L2, similarly to what was observed in the second experiment of Vaessen's (2017) study using a Rapid Serial Visual Presentation (RSVP) and Geipel et al.'s (2015) study. Geipel et al. (2015) observed no difference in occurrence rate between L1 and L2 and concluded that intuitive processes remain active in L1 and L2 during semantic illusion processing, opposed to the cognitive load hypothesis that supports the idea that an L2 promotes a switch from intuitive to more controlled processes (e.g., Costa et al., 2014; Keysar et al., 2012). Although Bautista Martín (2022) studied solely balanced bilinguals that acquired Spanish and Catalan early in their life, and thus a direct comparison of our results with those of Bautista Martín (2022) is not possible, it is worth to mention that Bautista Martín (2022) observed semantic illusions in both languages, as we did in the present study.

If we take into consideration the percentage of semantic illusions observed in previous studies, similar percentages were obtained for the sentences in the present research. For instance, Van Oostendorp and Kok (1990) observed a percentage of semantic illusions of approximately 32%. Similarly, Vaessen (2017), who collected data of Dutch-English bilinguals, showed a semantic illusions rate of roughly 30%. These percentages are very close to those observed in the present study (33% for monolingual native speakers, 33.1% and 43.1% for early and late bilinguals in GER respectively).

As we observed semantic illusions in L1 and L2, we confirm the postulates of the NST and related accounts to the semantic similarity hypothesis. Illusions occurred in both languages due to semantic similarities between critical words. As a consequence, theories seem to be more complementary than mutually exclusive. Indeed, results could also be described by the partial-match hypothesis or imperfect memory match hypothesis (Reder & Kusbit, 1991; see also Van Oostendorp & de Mul, 1990), given that the semantic overlapping between critical words may promote a partial match of the semantic information to stored mental representations (Dhaene et al., 2022, see also Reder & Kusbit, 1991), leading to semantic illusions. We recognize, however, that more research is needed in order to gain a more complete picture of the mechanisms underlying the effect of semantic illusions in other bilingual populations, speaking more or less closely related languages because linguistic transfer seems to be higher across languages with similar features (e.g., Spanish-Catalan bilinguals; Bautista Martín, 2022).

11. Conclusion

The main purpose of the present study was to examine the mechanisms leading to semantic illusions in bilingual processing by comparing early and late bilinguals. The above discussed findings for EP-GER bilinguals lead to the conclusion that no differences occur in the detection rate of semantic illusions in the L1 and L2, independently of AoA. These findings suggest that similar mechanisms underlie the occurrence of semantic illusions in monolingual and bilingual speakers. This may happen because intuitive processing leads to a partial match between the semantic information and the stored mental representations in the L1 and the L2 as complementary hypotheses hold, namely the partial-match hypothesis or the imperfect memory match hypothesis (Reder & Kusbit, 1991; see also Van Oostendorp & de Mul, 1990). Moreover, the results showing a similar percentage of semantic illusion in the L1 and L2 replicate those of previous studies (i.e., Geipel et al., 2015; Vaessen, 2017).

Acknowledgments. This study was conducted at the Psychology Research Centre (CIPsi/UM), School of Psychology and the Center of Humanities (CEHUM), School of Arts and Humanities, University of Minho, supported by the Foundation for Science and Technology (FCT) through the Portuguese State Budget (UIDP/01662/2020 and UID/00305/2020) and the Center of Linguistics (CLUP), the Department of Portuguese and Romance Studies (DEPER), University do Porto. The current study has also won the 'Prémio UMinho de Iniciação na Investigação Científica – 2021' for which we would like to thank the University of Minho.

Data availability statement. The code and data used in the experiments are available at: https://url.avanan.click/v2/___https://osf.io/a8vu6/?view_only=f69a25bf18674ab795bf7dfc22a4b4ea___.YXAxZTpjYW1icmlkZ2Vvcmc6YTpvOjMzYWM3MTFhMzRlYjFlNDczNzkyZTc0ODM1MjJhMmYwOjY6NjY2NjoyZjhjYWM0MDk1OTI4NzI4MTA3YzE0MzRiYTgzZTRlN2UzZTJjNTMyZGM3ZGY1Y2Q4ODhlZDVkZTc1MjY3MGRmOnQ6VDpG.

References

Alderson, J. C. (2005). *Diagnosing foreign language proficiency: The interface between learning and assessment*. A&C Black.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge University Press. https://doi.org/10.1017/CBO9780511801686

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Barton, S. B., & Sanford, A. J. (1993). A case study of anomaly detection: Shallow semantic processing and cohesion establishment. *Memory & Cognition*, 21(4), 477–487. https://doi.org/10.3758/BF03197179

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

Bautista Martín, A. M. (2022). *The bilingual moses illusion: Evidence for semantic illusions in Spanish-Catalan bilingual speakers* [Master's thesis]. University of Barcelona, Dipòsit Digital de la Universitat de Barcelona. http://hdl.handle.net/2445/190761

Blott, L. M., Rodd, J. M., Ferreira, F., & Warren, J. E. (2021). Recovery from misinterpretations during online sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 47(6), 968–997. https://doi.org/10.1037/xlm0000936

Bonfieni, M., Branigan, H. P., Pickering, M. J., & Sorace, A. (2019). Language experience modulates bilingual language control: The effect of proficiency, age of acquisition, and exposure on language switching. *Acta Psychologica*, 193, 160–170. https://doi.org/10.1016/j.actpsy.2018.11.004

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

Brysbaert, M., Buchmeier, M., Conrad, M., Jacobs, A. M., Bölte, J., & Böhl, A. (2011). The word frequency effect: A review of recent developments and implications for the choice of frequency estimates in German. *Experimental Psychology*, 58(5), 412–424. https://doi.org/10.1027/1618-3169/a000123

Costa, A., Foucart, A., Arnon, I., Aparici, M., & Apesteguia, J. (2014). "Piensa" twice: On the foreign language effect in decision making. *Cognition*, 130(2), 236–254. https://doi.org/10.1016/j.cognition.2013.11.010

Costa, A. S., Comesaña, M., & Soares, A. P. (2022). PHOR-in-One: A multilingual lexical database with PHonological, ORthographic and PHonographic word similarity estimates in four languages. *Behavior Research Methods*, 55(7), 3699–3725. https://doi.org/10.3758/s13428-022-01985-3

de Bruin, A., Carreiras, M., & Duñabeitia, J. A. (2017). The BEST dataset of language proficiency. *Frontiers in Psychology*, 8, Article 522. https://doi.org/10.3389/fpsyg.2017.00522

Dhaene, S., Dirix, N., Van Marcke, H., & Woumans, E. (2022). Moses or Noah? A case of 'potato-potahto'-when using a foreign language. *Bilingualism: Language and Cognition*, 25(3), 444–458. https://doi.org/10.1017/S1366728921000882

Dijkstra, T., Wahl, A., Buytenhuijs, F., Van Halem, N., Al-Jibouri, Z., De Korte, M., & Rekké, S. (2019). Multilink: A computational model for bilingual word recognition and word translation. *Bilingualism: Language and Cognition*, 22(4), 657–679. https://doi.org/10.1017/S1366728918000287

Erickson, T. D., & Mattson, M. E. (1981). From words to meaning: A semantic illusion. *Journal of Verbal Learning & Verbal Behavior*, 20(5), 540–551. https://doi.org/10.1016/S0022-5371(81)90165-1

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology*, 47(2), 164–203. https://doi.org/10.1016/S0010-0285(03)00005-7

Flores, C., Zhou, C., & Eira, C. (2022). "I no longer count in German." On dominance shift in returnee heritage speakers. *Applied Psycholinguistics*, 43(5), 1019–1043. https://doi.org/10.1017/S0142716422000261

García-Pentón, L., García, Y. F., Costello, B., Duñabeitia, J. A., & Carreiras, M. (2016). The neuroanatomy of bilingualism: How to turn a hazy view into the full picture. *Language, Cognition and Neuroscience*, 31(3), 303–327. https://doi.org/10.1080/23273798.2015.1068944

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

Gomes, J. N., & Maia, M. (2024). *Not noticing the evidence: How big are illusion effects on coordination, adjunction and embedding? Special Issue in: Recursion Across Languages – The Intricacies of Babel, Springer.*

Google. (2022). Google Forms. https://docs.google.com/forms

Green, P., & MacLeod, C. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, 7(4), 493–498. https://doi.org/10.1111/2041-210X.12504

Hannon, B. (2014). Research on semantic illusions tells us that there are multiple sources of misinformation. In D. N. Rapp & J. L. G. Braasch (Eds.), *Processing inaccurate information: Theoretical and applied perspectives from cognitive science and the educational sciences* (pp. 93–115). The MIT Press.

Huang, Y., & Ferreira, F. (2021). What causes lingering misinterpretations of garden-path sentences: Incorrect syntactic representations or fallible memory processes? *Journal of Memory and Language*, 121, Article 104288. https://doi.org/10.1016/j.jml.2021.104288

Jegerski, J. (2014). Self-paced reading. In J. Jegerski & B. VanPatten (Eds.), *Research methods in second language psycholinguistics* (pp. 20–49). Routledge.

Keysar, B., Hayakawa, S. L., & An, S. G. (2012). The foreign-language effect: Thinking in a foreign tongue reduces decision biases. *Psychological Science*, 23(6), 661–668. https://doi.org/10.1177/0956797611432178

Koornneef, A. W., & Van Berkum, J. J. A. (2006). On the use of verb-based implicit causality in sentence comprehension: Evidence from self-paced reading and eye tracking. *Journal of Memory and Language*, 54(4), 445–465. https://doi.org/10.1016/j.jml.2005.12.003

Koornneef, A., & Reuland, E. (2016). On the shallow processing (dis)advantage: Grammar and economy. *Frontiers in Psychology*, 7, Article 82. https://doi.org/10.3389/fpsyg.2016.00082

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Laufer, B., & Nation, P. (1999). A vocabulary-size test of controlled productive ability. *Language Testing*, 16 (1), 33–51. https://doi.org/10.1191/026553299672614616

Marian, V., Blumenfeld, H. K., and Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): assessing language profiles in bilinguals and multilinguals. *Journal of Speech Language and Hearing Research* 50(4), 940–967. https://doi.org/10.1044/1092-4388(2007/067)

Mitchell, D. C. (1984). An evaluation of subject-paced reading tasks and other methods for investigating immediate processes in reading. In D. Kieras & M. A. Just (Eds.), *New methods in reading comprehension research*. Erlbaum.

Muller, H., Resnik, P., & Phillips, C. (2020). *Explaining item-wise variability in Moses illusions. In 33rd annual CUNY conference on human sentence processing-Amherst, Massachusetts.*

Reder, L. M., & Kusbit, G. W. (1991). Locus of the moses illusion: Imperfect encoding, retrieval, or match? *Journal of Memory and Language*, 30(4), 385–406. https://doi.org/10.1016/0749-596X(91)90013-A

Sá-Leite, A. R., Flores, C., Eira, C., Haro, J., & Comesaña, M. (2023). Language balance rather than age of acquisition: A study on the cross-linguistic gender congruency effect in Portuguese–German bilinguals. *Bilingualism: Language and Cognition*, 26(5), 1079–1092. https://doi.org/10.1017/S1366728923000378

Shafto, M., & MacKay, D. G. (2000). The Moses, mega-Moses, and Armstrong illusions: Integrating language comprehension and semantic memory. *Psychological Science*, 11(5), 372–378. https://doi.org/10.1111/1467-9280.00273

Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition*, 128 (3), 302–319. https://doi.org/10.1016/j.cognition.2013.02.013

Soares, A. P., Machado, J., Costa, A., Iriarte, Á., Simões, A., de Almeida, J. J., Comesaña, M., & Perea, M. (2015). On the advantages of word frequency and contextual diversity measures extracted from subtitles: The case of Portuguese. *The Quarterly Journal of Experimental Psychology*, 68(4), 680–696. https://doi.org/10.1080/17470218.2014.964271

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

Treffers-Daller, J., & Korybski, T. (2016). Using lexical diversity measures to operationalise language dominance in bilinguals. In C. Silva-Corvalán & J. Treffers Daller (Eds.), *Language dominance in bilinguals: Issues of measurement and operationalization* (pp. 106–123). Cambridge University Press.

Vaessen, R. (2017). *The effect of semantic illusions on L1 and L2 sentence processing: Did Moses speak two languages on the Ark?* [Master's thesis]. Radboud University, Radboud Educational Repository. https://theses.ubn.ru.nl/handle/123456789/4935?locale-attribute=en

van Jaarsveld, H. J., Dijkstra, T., & Hermans, D. (1997). The detection of semantic illusions: Task-specific effects for similarity and position of distorted terms. *Psychological Research*, 59(4), 219–230. https://doi.org/10.1007/BF00439299

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

Van Oostendorp, H., & Kok, I. (1990). Failing to notice errors in sentences. *Language and Cognitive Processes*, 5(2), 105–113. https://doi.org/10.1080/01690969008402100

Zehr, J., & Schwarz, F. (2018). PennController for internet based experiments (IBEX). https://doi.org/10.17605/OSF.IO/MD832

Cite this article: Fernandes, A. I., Gomes, J. N., Haro, J., Flores, C., & Comesaña, M. (2024). The how and the when of semantic illusions in native and non-native languages, *Language and Cognition*, 1–25. https://doi.org/10.1017/langcog.2024.51